

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:51:20 ; Search time 1757.07 Seconds
(without alignments)
234.726 Million cell updates/sec

Title: US-09-396-196f-5

Perfect score: 25
Sequence: 1 gcgcagcaggtgcacgcgcagcatt 25

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size: 9
Total number of hits satisfying chosen parameters: 236099

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database: GenEmbl:*

1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_hlgo_hum:*
31: em_hlgo_inv:*
32: em_hlgo_rnd:*
33: em_hlg_hum:*
34: em_hlg_inv:*
35: em_hlg_rnd:*
36: em_hlg_other:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	6 AR029499	AR029499 Sequence
2	25	100.0	1041	6 AR034916	AR034916 Sequence
3	25	100.0	1084	6 A11530	A11530 B10B gene o
4	25	100.0	5793	1 EC0810	J04423 E.coli 7,8-
5	25	100.0	5872	6 A38246	A38246 Sequence 1
6	25	100.0	5872	6 A38251	A38251 Sequence 6
7	25	100.0	5872	6 A38251	A38251 Sequence 6
8	25	100.0	5872	6 A38251	A38251 Sequence 6
9	25	100.0	5872	6 A38251	A38251 Sequence 6
10	25	100.0	5872	6 A38251	A38251 Sequence 6
11	25	100.0	11022	1 AE000180	AE000180 Sequence
12	25	100.0	11501	1 AE005258	AE005258 Sequence
13	25	100.0	297816	1 AP002553	AP002553 Sequence
14	22	88.0	1121	6 E00893	E00893 Sequence
15	20	80.0	2923	1 MAV250020	MAV250020 Sequence
16	19	76.0	3603	1 PSEHRP1A	PSEHRP1A Sequence
17	19	76.0	14846	1 AE007050	AE007050 Sequence
18	18	72.0	44201	1 MTCY180	MTCY180 Sequence
19	18	72.0	37500	2 AC068068	AC068068 Sequence
20	18	72.0	38000	2 AC010079	AC010079 Sequence
21	18	72.0	42000	2 AC087836	AC087836 Sequence
22	17	68.0	12891	1 AE004192	AE004192 Sequence
23	16	64.0	10290	1 AR024397	AR024397 Sequence
24	16	64.0	14350	1 AE005685	AE005685 Sequence
25	16	64.0	14115	1 MTV031	MTV031 Sequence
26	16	64.0	14634	1 AE006923	AE006923 Sequence
27	16	64.0	23740	1 MTV026	MTV026 Sequence
28	16	64.0	34902	1 P01245436	P01245436 Sequence
29	16	64.0	49587	3 DMH48C10	DMH48C10 Sequence
30	16	64.0	60573	2 AC018267	AC018267 Sequence
31	16	64.0	82289	2 ATAC009895	ATAC009895 Sequence
32	16	64.0	115955	2 AC011732	AC011732 Sequence
33	16	64.0	177380	9 AC002457	AC002457 Sequence
34	16	64.0	180700	2 AC084136	AC084136 Sequence
35	16	64.0	289090	8 AE003424	AE003424 Sequence
36	16	64.0	1185	8 AF374832	AF374832 Sequence
37	15	60.0	3069	3 AF219383	AF219383 Sequence
38	15	60.0	3369	8 AF274671	AF274671 Sequence
39	15	60.0	3369	8 AF274671	AF274671 Sequence
40	15	60.0	11684	1 AE007551	AE007551 Sequence
41	15	60.0	15916	1 MTCY9CA	MTCY9CA Sequence
42	15	60.0	17988	1 AE007098	AE007098 Sequence
43	15	60.0	35209	1 AF065159	AF065159 Sequence
44	15	60.0	35573	3 AC009603	AC009603 Sequence
45	15	60.0	41193	2 AC013257	AC013257 Sequence
46	15	60.0	44545	2 AC010753	AC010753 Sequence
47	15	60.0	73691	2 AC014684	AC014684 Sequence
48	15	60.0	80332	2 AC010038	AC010038 Sequence
49	15	60.0	82852	2 AC017685	AC017685 Sequence
50	15	60.0	86217	2 HSDJ92032	HSDJ92032 Sequence
51	15	60.0	86254	2 AC013351	AC013351 Sequence
52	15	60.0	86896	2 HS93N13	HS93N13 Sequence
53	15	60.0	87814	8 AB046426	AB046426 Sequence
54	15	60.0	110000	2 LMFLCHR34_06	LMFLCHR34_06 Sequence
55	15	60.0	110000	2 LMFLCHR34_07	LMFLCHR34_07 Sequence
56	15	60.0	115915	2 AP000859	AP000859 Sequence
57	15	60.0	129949	2 AC022940	AC022940 Sequence
58	15	60.0	133707	2 AC087838	AC087838 Sequence
59	15	60.0	138965	3 AC003998	AC003998 Sequence
60	15	60.0	143805	9 HSDJ335E1	HSDJ335E1 Sequence
61	15	60.0	147223	2 AC090713	AC090713 Sequence
62	15	60.0	148038	2 AC013591	AC013591 Sequence
63	15	60.0	150391	2 AC011764	AC011764 Sequence
64	15	60.0	154306	2 AC034114	AC034114 Sequence
65	15	60.0	157376	2 AC010751	AC010751 Sequence
66	15	60.0	159930	2 AC014935	AC014935 Sequence
67	15	60.0	162345	3 AC006415	AC006415 Sequence
68	15	60.0	169524	2 AC023111	AC023111 Sequence
69	15	60.0	169524	2 AC023111	AC023111 Sequence
70	15	60.0	169524	2 AC023111	AC023111 Sequence

```

71 15 60.0 173681 2 AL353636
72 15 60.0 175643 2 AC018780 Homo sapi
73 15 60.0 176510 3 AC007054 Drosophila
74 15 60.0 180495 3 AC009847 Drosophila
75 15 60.0 181861 2 AC016322 Homo sapi
76 15 60.0 183937 2 AC037435 Mus muscu
77 15 60.0 185768 2 AP003971 Homo sapi
78 15 60.0 186002 3 AC006933 Drosophila
79 15 60.0 187235 2 AP001785 Homo sapi
80 15 60.0 187336 9 AL158822 Human DNA
81 15 60.0 194781 2 AC084279 Homo sapi
82 15 60.0 197303 9 AC012506 Homo sapi
83 15 60.0 199497 2 AC024534 Homo sapi
84 15 60.0 199607 2 AC009366 Drosophila
85 15 60.0 207842 2 AC010276 Homo sapi
86 15 60.0 209012 2 AC024989 Mus muscu
87 15 60.0 210956 2 AC090649 Mus muscu
88 15 60.0 280887 2 AE003526 Drosophila
89 15 60.0 295566 3 AE003597 Drosophila
90 15 60.0 296202 3 AE003785 Drosophila
91 15 60.0 298283 3 AE003785 Drosophila
92 14 56.0 221 3 AB002711 Entamoeba
93 14 56.0 282 11 G02917 human STS F
94 14 56.0 367 11 G24898 human STS F
95 14 56.0 373 10 MMWVA5TC human STS F
96 14 56.0 401 11 G29590 human STS S
97 14 56.0 479 11 G01446 Drosophila
98 14 56.0 499 10 BC002305 Mus muscu
99 14 56.0 574 2 AC031725 Giardia
100 14 56.0 606 2 AC056445 Giardia

```

ALIGNMENTS

```

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D,Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
    source          1..1041
                     Location/Qualifiers
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match          100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.004;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggcagcaggtgcatccagcatt 25
    |||
Db 79 GCGCAGCAGGTGTCATGCCAGCATT 103

```

```

RESULT 2
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5550521
KEYWORDS
SOURCE Unknown.

```

```

ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
    source          1..1041
                     Location/Qualifiers
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match          100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.004;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggcagcaggtgcatccagcatt 25
    |||
Db 79 GCGCAGCAGGTGTCATGCCAGCATT 103

```

```

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION B10b gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
AUTHORS
JOURNAL Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES
    source          1..1084
                     Location/Qualifiers
    gene            /organism="Escherichia coli"
    CDS             /db_xref="taxon:562"
                   24..1064
                   /gene="B10B"
                   24..1064
                   /gene="B10B"
                   /gene="B10B"
                   /codon_start=1
                   /transl_table=11
                   /protein_id="CAA00965.1"
                   /db_xref="GI:490218"
                   /db_xref="SWISS-PROT:P12996"
                   translation="MAHRPRWTLISVTELEFEKPLDLLFEAQVHROHFDPROVOST
                   LLSIKTGACPEDECKYCPDSSRYKTLGLEAEERLEMEVOVLESARKAKAGSTRFCGAM
                   KRHERDMPLYLEOMVOGVKAMGLEACMTLGTLSESOAOALNAGLDVYNNHLPDSPE
                   YGNITTRTYOERLDPLEKVRDAGIKKVGSGIVGIGEVKDRACLLIQALNLPDPES
                   VPINMLVKKVGTPLADNDVDVDFEIRITAVARIIMKPTSYRUSAGSEONNEQOAMC
                   FMAGANSIFYGCKLLTTPNPERDKLOLFRKILGINPOOTAVLAGDNDOORLROALMT
                   PTHDEVYNAAL"
BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

```

```

Query Match          100.0%; Score 25; DB 6; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.004;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggcagcaggtgcatccagcatt 25
    |||
Db 102 GCGCAGCAGGTGTCATGCCAGCATT 126

```

```

RESULT 4
LOCUS ECOBIO 5793 bp DNA
DEFINITION E.coli 7,8-diamino-7-phenylarsonic acid (bioA), biotin synthetase

```



```

(bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioC
protein, and dehydrobiotin synthetase (bioD), complete cds.
ACCESSION
J04423
VERSION
J04423.1 GI:145422
KEYWORDS
7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
bioC gene; bioD gene; bioF gene; biotin synthetase; dehydrobiotin
synthetase.
SOURCE
Escherichia coli (strain K-12) DNA.
ORGANISM
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia
1 (bases 1 to 5793)
Otsuka, A.J., Buoncristiani, M.R., Howard, P.K., Flamm, J. and
Johnson, O.
The Escherichia coli biotin biosynthetic enzyme sequences
predicted
J. Biol. Chem. 263, 19577-19585 (1988)
89066784
Draft entry and computer-readable sequence [1] kindly submitted by
A.Otsuka, 09-NOV-1988
Location/Oualifiers
1..5793
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
complement(98..574)
/note="ORF 1"
/codon_start=1
/transl_table=1
/protein_id="AA023513.1"
/db_xref="GI:1455168"
/translation="MKLISNDLRDGLKPHRHVFNKGNGYGDNIISPHLAMDVYPAQNK
SEVVCYDPDAPGSGMMHVVVNI.PADTRLPDGFSGGLVAMPDGLVQTRTDFGTG
VCGAPRKGTHRYLFTVHALDIERIDVEDGASAAVGFVNFHSLASITAMFS
complement(633..1925)
/gene="bioA"
complement(633..1925)
/gene="bioA"
/codon_start=1
/transl_table=1
/product="7,8-diamino-pelargonic acid aminotransferase"
/protein_id="AA023514.1"
/db_xref="GI:1457106"
/translation="MTTDLADPDRIHHPYMTSPPLVYVYSAEGCELLSDGR
LVDGMSMAAIIHGVNHLPOLNAKMSQIDAMSHVYFGCTTAPALIELCKLVAAGRN
ALECVFLADSGVAIVAMKALQYQAKGEAKORFLTFRNGYHGDTCGMSVCDPDN
SMHSLKGYLPENLFAPADPSRMDGEMEDRWGVARLMAARHETIAAVITETPOGA
GGRMYHPEWLRIRKICDREGILLADEIATFGRTGRLFACEHAELIAPILCLGKA
LNGCTMTSATLTTRVAETISNGEACFMHGPPTMGNPLACAAANASIALLESQDW
QVYADIEVQIREOLAPARDAEMVADVRYGIVGVEETHPVMAALOKFEVQGVIR
PEKLIYLMPTITLPOOLRLTAANRAVODETFECQ"
2012..3052
/gene="bioB"
2012..3052
/gene="bioB"
/codon_start=1
/transl_table=1
/product="biotin synthetase"
/protein_id="AA023515.1"
/db_xref="GI:145425"
/translation="MAHRPMTLSQVTELEKPLDLLEFAOVVHHQHPPOVYST
LLSTKTGACPEDCKYCPQTSRYKTGLAEALMEVEVLESARAKAGSTREMGAM
KNHPRDMPTLEOVGVKAMGLACMTLGLTESOARLANGLDYVHNHNDTSPES
YGNITTRTYBERLDLEKVDAGIKVSGIGLGTGKVDAGLLOJLANLPETPES
VPIINLVAKGTPLADNDDDADPFIPTIIVARIIMPTSYVRLSAGREOMEQOAMC
FMAGANSIEYGCCKLITTPNEDEKDLQFLFKGLINPOQTAVLAGDNDQOORLQALMT
PDTDEYNAAL"
3049..4203
/gene="bioF"
3049..4203
/gene="bioF"
/codon_start=1
/transl_table=1
/product="7-keto-8-amino pelargonic acid synthetase"
/protein_id="AA023516.1"
/db_xref="GI:145426"
/translation="MSWQEKINAAIDARRADALRRPYVPAOGAGRLVADROYLNF
SNDLIGLSHHPQIIIRAMOGAEOFGIGSGSVAGSYVYHQAELPEELAEMLGYSGRA
LLETSFANACVIAAMAKEDRIADRLSHASLEASUSPSLORFRAINDYTHLAR
LLASPGQOQVAVTEGVFSMDGSAPIAEIOQVQOHNQWMLMDVDAHGTGTEBOGRG
SCMLQKVPKEILLVTFEGKFGVGAACVCSSTVADYLLQFARHLIYSTSPPODAL
KSLAVIRSDGSDARREKLAALITPRBAGVODLPETLADSCSAIOPITVDNSHALOL
AEKLRQGCWVTAIRPVPYAGTALRLITLAAHEMQUIDLLVHLGNG"
4190..4945
/gene="bioC"
4190..4945
/gene="bioC"
/codon_start=1
/transl_table=1
/protein_id="AA023517.1"
/db_xref="GI:145427"
/translation="MATVNRQATIAAFGRAAHYEQHADLQROSADALLAMPORRYT
HVLDAAGCGPMGRHNRERHAYTALDLSPPMLVQROKQADADHYLAGDIESIPLATA
TFDLAENLAVOWCNGNLSTRELRYVRPKGVAFITLVQSLPERIQANQVDERP
HANFLPDEIEOSLNGVHYQHIIQITLTFPDGALAMSLKIGATHLHBSRPRIL
TRSOLORLQALAMPQOQGRPLTYHLFGLVARE"
4938..5597
/gene="bioD"
4938..5597
/gene="bioD"
/codon_start=1
/transl_table=1
/product="dehydrobiotin synthetase"
/protein_id="AA023518.1"
/db_xref="GI:145428"
/translation="MSKRTVYTGDTTEVGTQVASCALLQAKRAGRYTAGYKPVASGS
KATPEGLNSQALALQQRSSQLQDYATVNPYTPREPSPHISAOGRPIESLWMSAG
LRLPEQADQWALVDSAGMFTPLSDTETFDWYQEDLPYLIVGVKICGINAMTLA
QYIQAAGLTLGAVANDVTPPKRRAHEMTTLTRIPRCWERSPLQKIQAKROBES
"
BASE COUNT 1363 a 1554 c 1631 g 1245 t
ORIGIN 4626 bp upstream of HpaI site; 18 min on K-12 map.
Query Match 100.0%; Score 25; DB 1; Length 5793;
Best Local Similarity 100.0%; Pred. No. 0.0032;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggcgcagcagtgcatgcgcagcatt 25
Db 2090 gccgcagcagtgcatgcgcagcatt 2114
RESULT 5
A38246 5872 bp DNA PAT 05-MAR-1997
LOCUS A38246
DEFINITION Sequence 1 from Patent WO9408023.
ACCESSION A38246
VERSION A38246.1 GI:2294844
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia
1 (bases 1 to 5872)
Otsuka, A.J., Buoncristiani, M.R., Howard, P.K., Flamm, J. and
Johnson, O.
The Escherichia coli biotin biosynthetic enzyme sequences
predicted
J. Biol. Chem. 263, 19577-19585 (1988)
89066784
Draft entry and computer-readable sequence [1] kindly submitted by
A.Otsuka, 09-NOV-1988
Location/Oualifiers
1..5872
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
complement(98..574)
/note="ORF 1"
/codon_start=1
/transl_table=1
/protein_id="AA023513.1"
/db_xref="GI:1455168"
/translation="MKLISNDLRDGLKPHRHVFNKGNGYGDNIISPHLAMDVYPAQNK
SEVVCYDPDAPGSGMMHVVVNI.PADTRLPDGFSGGLVAMPDGLVQTRTDFGTG
VCGAPRKGTHRYLFTVHALDIERIDVEDGASAAVGFVNFHSLASITAMFS
complement(633..1925)
/gene="bioA"
complement(633..1925)
/gene="bioA"
/codon_start=1
/transl_table=1
/product="7,8-diamino-pelargonic acid aminotransferase"
/protein_id="AA023514.1"
/db_xref="GI:1457106"
/translation="MTTDLADPDRIHHPYMTSPPLVYVYSAEGCELLSDGR
LVDGMSMAAIIHGVNHLPOLNAKMSQIDAMSHVYFGCTTAPALIELCKLVAAGRN
ALECVFLADSGVAIVAMKALQYQAKGEAKORFLTFRNGYHGDTCGMSVCDPDN
SMHSLKGYLPENLFAPADPSRMDGEMEDRWGVARLMAARHETIAAVITETPOGA
GGRMYHPEWLRIRKICDREGILLADEIATFGRTGRLFACEHAELIAPILCLGKA
LNGCTMTSATLTTRVAETISNGEACFMHGPPTMGNPLACAAANASIALLESQDW
QVYADIEVQIREOLAPARDAEMVADVRYGIVGVEETHPVMAALOKFEVQGVIR
PEKLIYLMPTITLPOOLRLTAANRAVODETFECQ"
2012..3052
/gene="bioB"
2012..3052
/gene="bioB"
/codon_start=1
/transl_table=1
/product="biotin synthetase"
/protein_id="AA023515.1"
/db_xref="GI:145425"
/translation="MAHRPMTLSQVTELEKPLDLLEFAOVVHHQHPPOVYST
LLSTKTGACPEDCKYCPQTSRYKTGLAEALMEVEVLESARAKAGSTREMGAM
KNHPRDMPTLEOVGVKAMGLACMTLGLTESOARLANGLDYVHNHNDTSPES
YGNITTRTYBERLDLEKVDAGIKVSGIGLGTGKVDAGLLOJLANLPETPES
VPIINLVAKGTPLADNDDDADPFIPTIIVARIIMPTSYVRLSAGREOMEQOAMC
FMAGANSIEYGCCKLITTPNEDEKDLQFLFKGLINPOQTAVLAGDNDQOORLQALMT
PDTDEYNAAL"
3049..4203
/gene="bioF"
3049..4203
/gene="bioF"
/codon_start=1
/transl_table=1
/product="7-keto-8-amino pelargonic acid synthetase"
/protein_id="AA023516.1"
/db_xref="GI:145426"
/translation="MSWQEKINAAIDARRADALRRPYVPAOGAGRLVADROYLNF
SNDLIGLSHHPQIIIRAMOGAEOFGIGSGSVAGSYVYHQAELPEELAEMLGYSGRA
LLETSFANACVIAAMAKEDRIADRLSHASLEASUSPSLORFRAINDYTHLAR
LLASPGQOQVAVTEGVFSMDGSAPIAEIOQVQOHNQWMLMDVDAHGTGTEBOGRG
SCMLQKVPKEILLVTFEGKFGVGAACVCSSTVADYLLQFARHLIYSTSPPODAL
KSLAVIRSDGSDARREKLAALITPRBAGVODLPETLADSCSAIOPITVDNSHALOL
AEKLRQGCWVTAIRPVPYAGTALRLITLAAHEMQUIDLLVHLGNG"
4190..4945
/gene="bioC"
4190..4945
/gene="bioC"
/codon_start=1
/transl_table=1
/protein_id="AA023517.1"
/db_xref="GI:145427"
/translation="MATVNRQATIAAFGRAAHYEQHADLQROSADALLAMPORRYT
HVLDAAGCGPMGRHNRERHAYTALDLSPPMLVQROKQADADHYLAGDIESIPLATA
TFDLAENLAVOWCNGNLSTRELRYVRPKGVAFITLVQSLPERIQANQVDERP
HANFLPDEIEOSLNGVHYQHIIQITLTFPDGALAMSLKIGATHLHBSRPRIL
TRSOLORLQALAMPQOQGRPLTYHLFGLVARE"
4938..5597
/gene="bioD"
4938..5597
/gene="bioD"
/codon_start=1
/transl_table=1
/product="dehydrobiotin synthetase"
/protein_id="AA023518.1"
/db_xref="GI:145428"
/translation="MSKRTVYTGDTTEVGTQVASCALLQAKRAGRYTAGYKPVASGS
KATPEGLNSQALALQQRSSQLQDYATVNPYTPREPSPHISAOGRPIESLWMSAG
LRLPEQADQWALVDSAGMFTPLSDTETFDWYQEDLPYLIVGVKICGINAMTLA
QYIQAAGLTLGAVANDVTPPKRRAHEMTTLTRIPRCWERSPLQKIQAKROBES
"
REFERENCE
1 (bases 1 to 5872)
Otsuka, A.J., Buoncristiani, M.R., Howard, P.K., Flamm, J. and
Johnson, O.
The Escherichia coli biotin biosynthetic enzyme sequences
predicted
J. Biol. Chem. 263, 19577-19585 (1988)
89066784
Draft entry and computer-readable sequence [1] kindly submitted by
A.Otsuka, 09-NOV-1988
Location/Oualifiers
1..5872
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
complement(98..574)
/note="ORF 1"
/codon_start=1
/transl_table=1
/product="7,8-diamino-pelargonic acid aminotransferase"
/protein_id="AA023514.1"
/db_xref="GI:1457106"
/translation="MTTDLADPDRIHHPYMTSPPLVYVYSAEGCELLSDGR
LVDGMSMAAIIHGVNHLPOLNAKMSQIDAMSHVYFGCTTAPALIELCKLVAAGRN
ALECVFLADSGVAIVAMKALQYQAKGEAKORFLTFRNGYHGDTCGMSVCDPDN
SMHSLKGYLPENLFAPADPSRMDGEMEDRWGVARLMAARHETIAAVITETPOGA
GGRMYHPEWLRIRKICDREGILLADEIATFGRTGRLFACEHAELIAPILCLGKA
LNGCTMTSATLTTRVAETISNGEACFMHGPPTMGNPLACAAANASIALLESQDW
QVYADIEVQIREOLAPARDAEMVADVRYGIVGVEETHPVMAALOKFEVQGVIR
PEKLIYLMPTITLPOOLRLTAANRAVODETFECQ"
2012..3052
/gene="bioB"
2012..3052
/gene="bioB"
/codon_start=1
/transl_table=1
/product="biotin synthetase"
/protein_id="AA023515.1"
/db_xref="GI:145425"
/translation="MAHRPMTLSQVTELEKPLDLLEFAOVVHHQHPPOVYST
LLSTKTGACPEDCKYCPQTSRYKTGLAEALMEVEVLESARAKAGSTREMGAM
KNHPRDMPTLEOVGVKAMGLACMTLGLTESOARLANGLDYVHNHNDTSPES
YGNITTRTYBERLDLEKVDAGIKVSGIGLGTGKVDAGLLOJLANLPETPES
VPIINLVAKGTPLADNDDDADPFIPTIIVARIIMPTSYVRLSAGREOMEQOAMC
FMAGANSIEYGCCKLITTPNEDEKDLQFLFKGLINPOQTAVLAGDNDQOORLQALMT
PDTDEYNAAL"
3049..4203
/gene="bioF"
3049..4203
/gene="bioF"
/codon_start=1
/transl_table=1
/product="7-keto-8-amino pelargonic acid synthetase"
/protein_id="AA023516.1"
/db_xref="GI:145426"
/translation="MSWQEKINAAIDARRADALRRPYVPAOGAGRLVADROYLNF
SNDLIGLSHHPQIIIRAMOGAEOFGIGSGSVAGSYVYHQAELPEELAEMLGYSGRA
LLETSFANACVIAAMAKEDRIADRLSHASLEASUSPSLORFRAINDYTHLAR
LLASPGQOQVAVTEGVFSMDGSAPIAEIOQVQOHNQWMLMDVDAHGTGTEBOGRG
SCMLQKVPKEILLVTFEGKFGVGAACVCSSTVADYLLQFARHLIYSTSPPODAL
KSLAVIRSDGSDARREKLAALITPRBAGVODLPETLADSCSAIOPITVDNSHALOL
AEKLRQGCWVTAIRPVPYAGTALRLITLAAHEMQUIDLLVHLGNG"
4190..4945
/gene="bioC"
4190..4945
/gene="bioC"
/codon_start=1
/transl_table=1
/protein_id="AA023517.1"
/db_xref="GI:145427"
/translation="MATVNRQATIAAFGRAAHYEQHADLQROSADALLAMPORRYT
HVLDAAGCGPMGRHNRERHAYTALDLSPPMLVQROKQADADHYLAGDIESIPLATA
TFDLAENLAVOWCNGNLSTRELRYVRPKGVAFITLVQSLPERIQANQVDERP
HANFLPDEIEOSLNGVHYQHIIQITLTFPDGALAMSLKIGATHLHBSRPRIL
TRSOLORLQALAMPQOQGRPLTYHLFGLVARE"
4938..5597
/gene="bioD"
4938..5597
/gene="bioD"
/codon_start=1
/transl_table=1
/product="dehydrobiotin synthetase"
/protein_id="AA023518.1"
/db_xref="GI:145428"
/translation="MSKRTVYTGDTTEVGTQVASCALLQAKRAGRYTAGYKPVASGS
KATPEGLNSQALALQQRSSQLQDYATVNPYTPREPSPHISAOGRPIESLWMSAG
LRLPEQADQWALVDSAGMFTPLSDTETFDWYQEDLPYLIVGVKICGINAMTLA
QYIQAAGLTLGAVANDVTPPKRRAHEMTTLTRIPRCWERSPLQKIQAKROBES
"

```

Other publication FI 951547 950331
 Other publication JP 8501694T 960227.
 Location/Qualifiers
 1..5872
 /organism="Escherichia coli"
 /strain="DSM498"
 /db_xref="taxon:562"
 /clone="PB030A-15/9"
 1..96
 /function="PROMOTER PTAC"
 /evidence=experimental
 23..28
 /evidence=experimental
 45..49
 /standard_name="PROMOTER PTAC"
 /evidence=experimental
 105..119
 /standard_name="BIOB RBS NO.9"
 /evidence=experimental
 117..1157
 /gene="BIOB"
 117..1157
 /gene="BIOB"
 /codon_start=1
 /transl_table=11
 /number=1
 /evidence=experimental
 /product="BIOTIN SYNTHASE"
 /protein_id="CAA02325.1"
 /db_xref="GI:2294845"
 /translation="MAHRPRWTLSOVTELEFEPKPLIDLEFEAQVVRHOFDPPOVOST
 LLSIKTGACPECKTCPOSSRYKTGLFAERLMEVEYVLESARAKAAGSTRPCGAAM
 KPHFERDMFTLEOMVQVAKAGLEACMTLGLTSSQARLANAGLDYNNINLDTSP
 YXNITLTTRVQGLDLEKRYDAGIKVSGIYGLGTEVVRAGLILQNLTPRES
 VFINLVKGTGPIADNDVDAPFIRITAVARIMPTSVRLSAGEQNNEDTOAC
 EMAGNSIFYGCKLITTPNPECKDLQLEFKLGINPOTAVLAGDNEQDRLQALMT
 PPTDEYNAAL"
 2284..2297
 /standard_name="BIOC RBS"
 2295..3050
 /gene="BIOC"
 2295..3050
 /gene="BIOC"
 /function="INVOLVED IN PIMELYL-COA SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /number=3
 /product="PROTEIN"
 /protein_id="CAA02326.1"
 /db_xref="GI:2294846"
 /translation="MATVNRKQATIAAFGRAAHYEQHADLQROSDALLAMTPQKKT
 TFDLAWGCGPMGRHREHRAQVYALDLSPPMVYOROKDADADHYLAGDTESPLATA
 HANRFLPDEIEOSLNGVYOHIOPIITIMPDALSAMRSLKIGATHLHEGRDRIIL
 TRSOLRLQALWQGGRYPLTYHLGVATHE"
 3742..3752
 /standard_name="BIOA RBS"
 3750..5039
 /gene="BIOA"
 3750..5039
 /gene="BIOA"
 /standard_name="S-ADENOSYL-L-METHIONINE:
 8-AMINO-7-OXONONANICATE AMINOTRANSF.""
 /ec_number="2.6.1.62"
 /codon_start=1
 /transl_table=11
 /number=5
 /evidence=experimental
 /product="DAPA SYNTHASE"
 /protein_id="CAA02327.1"
 /db_xref="GI:2294847"
 /translation="MTTDDLAFDQRHIMHPYTSMTSPPLYPVVSABGCELLSDGR
 LVDSGSMWAAIHGNIHNPQINAAKRSQIDAMSHVFGGITHAPAIELCRKLVAMPQD
 LECVFLADSGSAVEAVAKMALQYQAKGEARQRFLEFRNGYHGDTEFGANSVCDPDNS

MSLMKGYLLENLFPAPQSRMGCEWDERDMGFAFLMAHHHEIAAVIIEPIYOGAG
 GMRVYHEPHEMLKRIKICDREGILLIADETATGGRGKLPACENHAIADILGKAL
 TGGTMTLSATLITREVAETLSNEACGFHNGPTFMKNPILACAAANASIALIESGMOO
 QVADIEVQLEQOLPARDAEMVADVIVLAIQVETTHPVYMAALQKFFVEGVIIRP
 FGKILTYMPPIYILLPOOLRLTFAVNRVAVODETFPCQ"
 5088..5100
 /standard_name="ORF1 RBS"
 5098..5574
 /gene="ORF1"
 5098..5574
 /gene="ORF1"
 /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /number=6
 /evidence=experimental
 /product="PROTEIN"
 /protein_id="CAA02328.1"
 /db_xref="GI:2294848"
 /translation="MKLISNDLRDQDKLPHRRVFNMGSGYDGNISPLAMDVYPAGTK
 SFVVTGYDPAFTGSGMHWVYNLPADTRVLPGQSGIYAMPDGYLOTRTDPGKTG
 YGGAAPKPGETIRYIFTVHALDIERIDYDEGASGAMGFVNRHSIASISITAMFS"
 5583..5605
 /standard_name="RHO-INDEPENDENT TRANSCRIPTIOMAL
 TERMINATOR"
 5583..5644
 /standard_name="RHO-INDEPENDENT TRANSCRIPTIOMAL
 TERMINATOR"
 BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN
 Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No 0.0032;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 qcgcgagcgatgcacgcgcagcat 25
 |||||||||||||||||||||||||
 Db 195 GCGCAGCAGGTCATCGCCAGCAT 219
 RESULT 6
 A38251 5872 bp DNA PAT 05-MAR-1997
 LOCUS
 DEFINITION
 A38251
 ACCESSION
 A38251.1 GI:2294849
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli.
 Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 1 (bases 1 to 5872)
 BIRCH.O., Brass, J., Fuhrmann, M. and Shaw, N.
 BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
 Patent: WO 9408023-A 6 14-Apr-1994;
 IONZA AG (CH)
 Other publication PL 308301 950724
 Other publication CA 2145400 940414
 Other publication AU 4820293 940426
 Other publication HU 71781 960228
 Other publication SK 42095 951108
 Other publication CZ 9500809 950913
 Other publication FI 951547 950331
 Other publication JP 8501694T 960227.
 Location/Qualifiers
 1..5872
 /organism="Escherichia coli"
 /strain="DSM498"
 /db_xref="taxon:562"
 /clone="PB030A15-9"
 1141..1156
 /standard_name="BIOF RBS"
 1154..2308
 /gene="BIOF"

/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
 /codon_start=1
 /transl_table=1
 /evidence=experimental
 /product="PROTEIN"
 /protein_id="CAB69593.1"
 /db_xref="GI:6741866"
 /translation="MKRLSNDLRGDKLPHRVHNGMGYDGNISPHLAWDDVPA;T K
 SFVYTCYDPAPPGSSGMHWVNALPADTRVLPGFGSGSLVAMPDGLQTRDFGKIG
 YDGAAPPKGGEHRIFTVHALDIERIDVEGASGAVGVNFHSLASASTVAMS"
 5583..5605
 5583..5644
 /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
 TERMINATOR"

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.0032;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccgacagatgcacgccagcatt 25
 ||||||||||||||||||||
 Db 195 GCCGACAGCGTGCATGCCACGACATT 219

RESULT 8
 A93679 5872 bp DNA PAT 22-JAN-2000
 LOCUS A93679
 DEFINITION Sequence 6 from patent EP0798384.
 ACCESSION A93679
 VERSION A93679.1 GI:6741867
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., and Brass, J.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
 LONZA AG (CH)

FEATURES
 source Location/Qualifiers
 1..5872
 /organism="Escherichia coli"
 /strain="DSM498"
 /db_xref="taxon:562"
 /clone="PRO30A15-9"
 1141..1156
 /standard_name="BIOD RBS"
 1154..2308
 /gene="BIOP"
 1154..2308
 /gene="BIOP"
 /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
 /EC_number="2.3.1.47"
 /transl_table=1
 /codon_start=1
 /evidence=experimental
 /product="KAPA SYNTHASE"
 /protein_id="CAB69594.1"
 /db_xref="GI:6741868"

translation="MSWQEKINAALDARRADALRRYPVAGAGRMVLVADDDQYLNE
 SSNDYLGISHPOIIRAMQCAEFGISGSGSHVSGYSVHQALEFLEAWICYSRA
 LIFSGFANQAVIAAMAKEDRIADRLSHASILEASISPSQLRFAHNDVTHLAR
 ILASPCQGMVTEGVSGMDGSAPIAEIOYQOINGMVDDAGTGVIEQGRG
 SCMLQKVPDELLVTFEGKPGVSCAAVICSSTVADYLIQFARHLIYSTSWPAAQAL
 RASLAVRSDESDARREKLAITFRAGVODLPTLADSCSAIOLITVDNSRALOL
 AEKLEQCCWTAIPPTVPAGTARLRLTLTAHEMODIDRLLEVLHGNC"
 3030..3045
 /standard_name="BIOD RBS"
 3043..3753

RBS
 gene

CDS
 /gene="BIOD"
 3043..3753
 /gene="BIOD"
 /standard_name="DETHIOBIOTIN SYNTHASE"
 /EC_number="6.3.3.3"
 /codon_start=1
 /transl_table=1
 /evidence=experimental
 /product="DTB SYNTHASE"
 /protein_id="CAB69595.1"
 /db_xref="GI:6741869"

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.0032;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccgacagatgcacgccagcatt 25
 ||||||||||||||||||||
 Db 195 GCCGACAGCGTGCATGCCACGACATT 219

RESULT 9
 AR101809 5872 bp DNA PAT 14-FEB-2001
 LOCUS AR101809
 DEFINITION Sequence 1 from patent US 6083712.
 ACCESSION AR101809
 VERSION AR101809.1 GI:12812607
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M., and Shaw, N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
 Location/Qualifiers
 1..5872
 /organism="unknown"
 BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.0032;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccgacagatgcacgccagcatt 25
 ||||||||||||||||||||
 Db 195 GCCGACAGCGTGCATGCCACGACATT 219

RESULT 10
 AR101810 5872 bp DNA PAT 14-FEB-2001
 LOCUS AR101810
 DEFINITION Sequence 6 from patent US 6083712.
 ACCESSION AR101810
 VERSION AR101810.1 GI:12812608
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M., and Shaw, N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 6 04-JUL-2000;


```

/protein_id="AAC73861.1"
/db_xref="GI:1786991"
/translation="MTTDDLAFDQRIHHPYTSMTSLPVVSAEGCELISDGR
LVDGSMMAALHGTHNPQLNAAMKSQIDAMSHVFGITIPAPAEICRKIVAMP
LECVLADSGSVAVVAVAMKALQYWAQAGARORPLTFRNGYHGDITGAMSCD
MHSIMKGYLPENLFAPOSRMDEMDRAGVAFRLMAHRIETIAVITTEPVOCG
GMRYHPEMLKRIKICDREGILLIADITANGPCTGKIFACPAEELAPDLICGK
TGTMTLSATLTREAVETISNGEACGMHGPTEGMNPLACAAANASLALLES
QVADIEVLEQOLAPARDAEVADVRVIGALGVETTHPVMALQKFEQGVWIRP
FGKILYMPYITIPQOLORTIAAVRAVDETFEQ"
complement(2193..2221)
/gene="biola"
/note="Factor Sigma70; predicted +1 start at 807260"
3411..3450
/note="central position to biob promoter: -20"
3411..3450
/note="central position to predicted promoter: -20; genetic
evidence for the site"
/bound_molety="Biob predicted site"
complement(3411..3450)
/note="central position to predicted promoter:50"
/bound_molety="Biob predicted site"
complement(3411..3450)
/note="central position to biob promoter:50"
3413..3441
/bound_molety="Biob documented site"
3413..3441
/note="Factor Sigma70; promoter biob; documented +1 at
808525"
complement(3447..3473)
/note="Factor Sigma70; promoter biob; documented +1 at
808515"
3493..4533
/gene="biob"
/note="b0775"
3493..4533
/gene="biob"
/EC_number="2.8.1.-"
/function="enzyme: Biosynthesis of cofactors, carriers:
Biotin"
/note="0346: 99 pct identical to BI0B_ECOLI SW: P12996"
/codon_start=1
/translation="biotin synthesis, sulfur insertion?"
/protein_id="AAC73862.1"
/db_xref="GI:1786992"
/translation="MAHPRMTLSQVTELEKPLDLLEAQQVHRQHEDPROVST
LSTKIGACPEDCYKCPQSSRYKGLLEAEERMEYVLESARAKAAGSTRECGAAW
KNPHERDMPELVQWQVAVKAMGLEACMTLGLTSEAOARLANAGIDYYNNHLDSP
VGNITITRYOERLDTLEKVDAGIKVCSGGIVGKXDRAGILLIOLANLPTPES
VFINMLVYKGTPLADNDVDAPDFIRITIAVARIMPSTYVRLSGRQOMEOQAMC
PMDEYVNAAL"
4530..5684
/gene="biob"
/note="b0776"
4530..5684
/gene="biob"
/EC_number="2.3.1.47"
/function="enzyme: Biosynthesis of cofactors, carriers:
Biotin"
/note="0384: 100 pct identical to BIOF_ECOLI SW: P12998"
/codon_start=1
/translation="1"
/product="8-amino-7-oxononanoate synthase"
/protein_id="AAC73863.1"
/db_xref="GI:1786993"
/translation="MSWQEKINAALDARRADALRRYPVAGAGRWLVADROVYNF
SSNDYGLSHHPQITRAMQGAQGFQSGSGSVSYVHGLEELAEELAEWYGSRA
LLETSGFANAQVIAAMAKEDRIADRLSHASLLEAASLSPSLRFFAHNYTHLAR
LASPCGQOMVTEGVSMDSAPLAEIOQVTOOHNGWLMWADHAGCYVCEQKRG
SCMLQKVPPELLVTEFGKGVGSAVILCSSTVADYILLOFARHLITSTMPPAQAL"

```

```

Query Match      100.0%; Score 25; DB 1; Length 11022;
Best Local Similarity 100.0%; Pred. No. 0.0029;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggcagcagatgcatcgccagcatt 25
Db 3571 GCGCAGCAGCGTCGCCAGCATT 3595

RESULT 12
AE005258
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
1 (bases 1 to 13501)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grobeck,E.J., Davis,N.W., Lim,A., Dimantanta,E., Potamouzis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
11206551
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
2 (bases 1 to 13501)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grobeck,E.J., Davis,N.W., Lim,A., Dimantanta,E., Potamouzis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1..13501
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohemorrhagic"
<1..7576
/note="O-island #36; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655; Cryptic prophage
CP-933K; Includes one copy of the 13 bp direct repeat that
flanks the prophage"
66..665
/gene="lomk"
/note="20981"
66..665
/gene="lomk"
/function="putative membrane; Other or unknown (Phage or
Prophage Related)"
/note="Residues 1 to 199 of 199 are 72.86 pct identical to
g117532789|gb|AAFG3231.1|AF151091_2 (AF151091) lom
[prophage P-EBD]"
/codon_start=1
/translation="1"
/product="putative outer membrane protein of prophage
CP-933K"
/protein_id="AAG55137.1"
/db_xref="GI:12513752"
/translation="MRKYCAILSAICIAVSGVPANASHQSTLSAGYLAHSTADAG
SDDLNGINVKRYRETTDLGLITSFSTANAEDEQKTHYSPTRMHEDYVRNKFVSVMAG

```

```

gene
CDS
/psrvnwmfmsaymagvayvrvyfpfsgdyfrvtdnkrrkthdvlrgsdarysnstslaw
gagvofnpfsvaavayvayexsgsgdmrdtgdgfvgykpf"
719..2041
/gene="20982"
719..2041
/function="putative structure; Structural component (Phage
or Prophage Related)"
/note="Residues 164 to 440 of 440 are 68.79 pct identical
to residues 381 to 645 of 645 from Genpept 118 :
g1145854361gb|AAD25464.1|AF125520.59 (AF125520) putative
tail fiber protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="putative tail component of prophage CP-933K"
/db_xref="GI:12513753"
/transl_table="MGMMAVOISGVLCDCAGKPIONCTIOLKARRSTXVVVNTVASE
MPDAGRISMDVEYGOYVTLVEGFPSPHAGTISYEDSOQGTINDPLGAMIPEDAR
PEARLRFEQWVEEAHAEAKKAGEAETISARNAGTSASKASANAADTSADASE
SAQOAAESAASAKSEASASSASAQAASJOSATPAELSKTAESAAGAAADA
TSTEKARESAESAQESORIAEDAVNR1PTVGPDPGKPGPGAGPGGQDGE
KQDTGAGATGGERGGDTPGAGPGKGRGERGELTGNAGPGGPGKGDGAAPPA
PGPGKGTGAACPVATGPGPGKGRGERGELTGNAGPGGPGKGDGAAPPA
TGLTFDPKDATQVQGFHLYVRFQDGPQDVKGLDEVSQDTRTGE"
2004..2312
/gene="20984"
2004..2312
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 1 to 102 of 102 are 98.03 pct identical to
residues 28 to 129 of 129 from Genpept 118 :
g1145854371gb|AAD25465.1|AF125520_60 (AF125520)
hypothetical protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55139.1"
/db_xref="GI:12513754"
/transl_table="MKMAVIOAELENDNMILRKIMSLCCGKHDCENRSLTAOLR
LGPADLESDENGILPEODRVTQVVLADAKQIQCVVRPLQILRADGTWENTGKMK
"
2489..3469
/gene="20985"
2489..3469
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 13 to 326 of 326 are 55.73 pct identical
to residues 18 to 331 of 336 from Genpept 118 :
g1169603671gb|AAF33527.1 (AF170176) hypothetical protein
predicted by Glimmer [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55140.1"
/db_xref="GI:12513755"
/transl_table="MLSPRTTFHNSVNIYVSSPCOTVSFAGKEVELKVIDEKTPIILF
OMFEPNPERKKDEVPIVNTKQHPYLVDVNTAKRESIRMGICIPDQGFSPNOKTAPS
KLEFDPENVM1IYREDVDSMTDKLSDIYHDIICEORLKRDEKDEYLLMLEELK
RTSKAODSLISMYAKRNHAWDFPRLALAKAGFIFCTYNTKNHG1SFEQCIYLD
MMLITLGLGTIYADGATSMHVDNRDSDNIENSA1IVNRSNHPALLEGLSFMSKVD
AHRYVDGLGKGVKKYFNFTPLHNYNHPCDFLEFNINPTIMNTSQYTCSSW"
3530..4522
/gene="20986"
3530..4522
/gene="20986"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11

```

```

/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55141.1"
/db_xref="GI:12513756"
/transl_table="MKIPSLQSNFNSAPAGYSNPIAPRANAENAVADYLDIGKR1PL
SAADLSNVESEVIRAVHDSRSK1DQHTVMTIGTLYDALSRQOTRDAVSYGIHNEK
VHIGCIKYRMEYELNESSVK1DIDQSLTNELEYEDVGEPIFPICEGENENEBRY
VSFVAPDPTDSEMPQWQEG1HELIHHVTVSGSPGSDSN1ELGPELILARRAOELG
WSVPDGPVAGAREARERLRLNMLARQAAHREHNERAFERLGTISDRYEAASPDPT
EYSVSNIGYGF1QOHDFPGLAINDN1QDANQ1QDLYGARYIFTFGDVKNHQO"
5350..6231
/gene="20989"
5350..6231
/gene="20989"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55142.1"
/db_xref="GI:12513757"
/transl_table="MLSPYSVNIQGSNLSLRNLTPDNRVLSVRDAVHSDGAY
KGNRTYRVVATDNKFCYTRSHSCFTNLHLRLQMPKGEISRIEVLVLAAPSAM
EKGIVHSNRPDLPVDVAPPELPSYDRIKLSVPGVIGKGNNAVYEDAPATVLMK
FTTSQSENEVTSVRCFNOYVAGSAEK1YGNNGD1IGIRMDK1NGESLNLSSLPQO
AHHAIYDMDRLEOKGILFVDTETNVLVYDRAKNFNP1DIDISYVNSDRWSQIMQ
SYHGKODDISVLSKI"
6462..7160
/gene="20990"
6462..7160
/gene="20990"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55143.1"
/db_xref="GI:12513758"
/transl_table="MRPTSLNVLHQSSRSSMSPDIESLKVASSVQIKNNPOLRF
OGTDHNI1YOOIEALDLKISTETGRVLAISISLKSSEYV1HILNSRLGVMAHD
IDAENHGRGSDPCHNLNAVEYPCGEGISVDFPHAT1YFHELHAFHNLNGRLKVES
SRAESQKSP1LLEEARVGLAFAFSEVSEKMFHEIGMPRTSYXPXSALIHDDNT
VSLGFOQVRLHPL"
complement(7668..8144)
/gene="ydhB"
/note="ydhB"
complement(7668..8144)
/gene="ydhB"
/function="orf; Unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 strain
MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/transl_table="MKLISNDLSDGDKLPRHRHFNNGMDYDGNISPHLANDDYPAKTX
SFVVTCYDPAPTGSGMWVNVNLPADTFVLRLQMGSGGV1VAMPDVLQTRDQFRTG
YDGAAPRGKHRYIFTHALDVERIDVDEGASGAMVFNVRHSHLASISITAMFS"
complement(8203..9492)
/gene="ydhA"
/note="ydhA"
complement(8203..9492)

```

```

Query Match      100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 0.0028;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1  ggcacagcaglgacatgccagcatt 25
DB      9657  gccagcagcagtcattgccagcatt 9681

```

```

RESULT 13
LOCUS AP002553 297816 bp DNA BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION AP002553 BA000007
VERSION AP002553.1 GI:13360211
KEYWORDS
SOURCE
ORANISM
DNA. Escherichia coli O157:H7 (strain:O157:H7, sub-strain:RIMD 0509952)
REFERENCE
AUTHORS Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Yanamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C. and Shinagawa,H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
Genes Genet. Syst. 74 (5), 227-239 (1999)
TITLE
JOURNAL Genet. Syst. 74 (5), 227-239 (1999)
MEDLINE 20198780
REFERENCE
AUTHORS
TITLE 2 (sites)
Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
Comparative analysis of the whole set of RNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
JOURNAL 20557356
MEDLINE
REFERENCE
AUTHORS 3 (sites)
Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yanamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
Shinagawa,H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
JOURNAL 20564182
MEDLINE
REFERENCE
AUTHORS 4 (sites)
Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shinagawa,H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
JOURNAL 21156231
MEDLINE
REFERENCE
AUTHORS 5 (bases 1 to 297816)
Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
Hayashi,T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
URL:ftp://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
Fax:81-6-6879-2047)
COMMENT genome project.
FEATURES
location/Qualifiers
1..297816
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
79..1245
/gene="ECs0753"
79..1245
/gene="ECs0753"
79..1245
/gene="ECs0753"

```

```

/note="similar to SUCC_ECOLI g11786949 percent identity
100 in 388 aa (Conserved in E.coli K-12)"
/transl_table=11
/evidence=not_experimental
/product="succinyl-CoA synthetase beta subunit"
/protein_id="BAB34176.1"
/db_xref="GI:13360212"
/translation="MNLHEYOKAFARVGLPAPVYACTTPRAEFAASTIGAGPW
VKOVHAGGRGAGGVKYNKEDIRARENKLRVLTOTDANGPVDNILEAAT
DIARELYGAVYDRSSRRVYRNASTEGVEIEKVAEETPHLIHWVALDPLGPPYOG
RELAKRLGLEGLVVOEFTKEGLATFLEERLALIEINPVTIKOGDLIDCKIGA
DGNALFROPDLREMRDQSOEDPREAOAQMELNVAALDGNIGCMWAGLGMMDIY
KLHGEPAFNLVGGGATKERVTEAFKITLSDPKKAVLVNIFGIVYICDLIAGIIG
AAVEGVNVPVVRLEGNNAFLGAKKLADSGLNTIAKGLDAAQYVAANVEG"
1245..2114
/gene="ECs0754"
1245..2114
/gene="ECs0754"
/note="similar to SUCC_ECOLI g11786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/translation="MSILIDKNTKVICOGFTSGQTFHSEQATVATGKMGVYPPG
GTHLGLFVFTVEAVATGATSVIVYVAPFCKDSILTAIDAGIKLITITTEGTP
LDMLTGVKRLDEAGVVMGPMPCGVITTPGKRGIGTIOGHIRKPKGVIVSRGTLVE
AAVKTQDYGCSQSCVYIGGDPITGNSNFTLLEFEKDPQTEAIVMIGETIGSAEEA
AAYIKENYTRKPVYVITAGVTPAKGRMHGAGAITAGKTADEKFALEAAGVKTAS
LADIGALKTVLK"
2781..3686
/gene="ECs0755"
2781..3686
/gene="ECs0755"
/note="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator) -
Escherichia coli g1417043 [sp|P32064|GCVA_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/translation="MRGKIPKTELVTVEEVVARESYTRAERLATOSAVROVSAL
EEFLRPLFSHKKRIEFLNDTGKYYGIYKETLKEPNTIMTQPTVOYIELAVN
PTFTSHWLINLHEFTKIPHDIVINHSIANNKQDPLNREYDAIKRENCATVAEYV
LFEFEITLPCVSSGLIANSODKLSVAELTELPLILHSTSTITGWEEMFALGVSSPLVN
NKPREDLSMLIAVNSNIGVALPRFAIOHDDSDMWIPCDVPIRTGNRFRIMTQW
EKSDDPHLDQFRFWMFLAKSVYVQEM"
complement (3720..4322)
/gene="ECs0756"
complement (3720..4322)
/gene="ECs0756"
/note="probable cob(II)amin adenosyltransferase, similar
to cob(II)amin adenosyltransferases (corrinoid
adenosyltransferases) e.g. [Escherichia coli]
g1115148 [sp|P13040|BRUR_ECOLI percent identity 67 in 200
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative cob(II)amin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translation="MEARISTEERHVOOROKLEQVDYTRVAAIEKKGLTIYFTGNGK
KSTAFGTVTRAVGKTVGYAOYIKGMDNGCEVNLQPLGCFRIMNGTGTWENRK
QADIDAKFVWSESKRMLADKRYDILVIGDELTYMLAIYHILDTVEVYIASLQNRPAQSY
IYIGRCHSQILKMDYVSEIRVKAHFDNGIOAPGIDW"

```



```

gene      complement(4332..5984)
          /gene="ECs0757"
          /complement(4332..5984)
          /note="probable fumarate hydratase, similar to fumarate
          hydratases - e.g. fumarate hydratase class I, aerobic
          (fumarate) - Escherichia coli
          g11205981sp1P009231F0MA_EC001 percent identity 68 in 545
          aa"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="putative fumarate hydratase"
          /protein_id="BAB34180.1"
          /db_xref="GI:13360216"
          /translation="MSKPFIMQEPFLQNKDCTEYTLISDQHTVTELDGEVYKIAPE
          ALTLARQAFYEASFLRSALHLOVASILINDQASNDKRYVALQLRANAEVSKGLP
          NCDQGTATVASKGOIWTGNDALSKGIYSTPOENKLRSQANAPLDMYTEVNTQ
          TNLPAQIDISAVAGDEYHPLCVNKGGSANKALVQETSLQPEKTLFLIEKMSL
          GTAACPPYHIAFVVGGLSDOTLTKAKLASTKTYDMLPTSGNCGCAFEDIEKYL
          EASQDFGIGAOFQGYFAHDIVIRLPRHGSGCIAMALSCSADRNIKAKINHGTL
          EKLEHNGQYTPASLRBNHQAQVOLDLNRPLDVMODLARLPVGTNRSLSGTVAR
          DIAHAKIKARLDSEPMPEYELKHIIYVYAGPAKTPENMACGSLGPTTGKMDGYIDTF
          OAAAGSLVMSLSTKNSQOVTDACRKHGGFNLSIGGAALLAGEVYKSLRCEYPELG
          MEAVMMMEVENVLPATLVVDKGNPFSSQFEQHRCASCPAGH"
          /complement(6092..7372)
          /gene="ECs0758"
          /complement(6092..7372)
          /note="probable transport protein, similar to
          glutamate/aspartate transport proteins (proton glutamate
          symport proteins) e.g. [Bacillus sterothromophilus]
          g112124671sp1P249431GLTT_BACST percent identity 38 in 416
          aa, also similar to Ca-dicarboxylate transport"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="putative transport protein"
          /protein_id="BAB34181.1"
          /db_xref="GI:13360217"
          /translation="MKKISLTTMLLALVGMIIIGVNLNTASPEFKKLYAOEISIT
          TIFRLIKMTIAPLVSVTLVGIKMDAKLRIGESKTLFLFCASLSIALGLITV
          NFMVGTGINFVAHAETGVAAEPFLKVFISHAPTSIVDMANNEILQIVFSI
          FLGSLAIGEKSAIYHADBLSLAHMLKLTGYMLPALTFAIISALLAEKRLAA
          VSACIFMGEFEFTMLLMVLLIGLAIYVGCIRLRALSEPALLAFTSSSAAP
          GTLKLEDFGVSPKIASRVLPICGSPNLSGMACSFATVIAQACNIIHISIGROITM
          LLIIMLTSKMGAVPRASMVVIAATLNQENIPEAGLITLLMGVDFPLMGASATVMSN
          AMGAAMVSRMGEHEGECRCGKALKPESNVALP"
          /complement(7533..7850)
          /gene="ECs0759"
          /complement(7533..7850)
          /gene="ECs0759"
          /note="unknown"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="hypothetical protein"
          /protein_id="BAB34182.1"

Query Match      100.0%; Score 25; DB 1; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1  ggcgcagcagtgatgcgcagcatt 25
          |||||||||||||||||||||||||
Db      92389  GCGCAGCAGGTGCATCGCCAGCATTT 92413

RESULT 14
LOCUS      E00893      1121 bp      DNA      PAT      29-SEP-1997
DEFINITION      Genomic DNA encoding biotin Synthetase.
ACCESSION      E00893

```

```

VERSION      E00893.1      GI:2169154
KEYWORDS      JP 1986149091-A/1.
SOURCE
ORGANISM      Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 1121)
Hirono, Y., Kojima, T. and Kimura, H.
PUBLICATION      DUPLICATION OF BIOTIN
PATENT: JP 1986149091-A 1 07-JUL-1986;
NIPPON SODA CO LTD
OS      Escherichia coli
PN      JP 1986149091-A/1
PD      07-JUL-1986
PE      24-DEC-1984      JP 1984272605
PT      HIRONO YOSHIOHKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
C12N15/00, C12N1/20, C12P13/18, (C12N1/20, C12P13/18, C12P13/18, PC
C12P13/19);
CC      strandedness: Double;
CC      topology: linear;
CC      hypothetical: No;
CC      anti-sense: No;
CC      *source: Strain=Escherichia coli Nsl01;
CC      Feature is identified by experimental;
FH      key      Location/Qualifiers
FT      CDS      42..1079
          Location/Qualifiers
          1..1121
          /organism="Escherichia coli"
          /db_xref="taxon:562"
          289 a      296 c      325 g      211 t

BASE COUNT      289 a      296 c      325 g      211 t
ORIGIN

Query Match      88.0%; Score 22; DB 6; Length 1121;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      4  cagcagatgcacgcgcagcatt 25
          |||||||||||||||||||||||||
Db      123  CAGCAGGCGATCGCCAGCATTT 144

RESULT 15
LOCUS      MAV250020      2923 bp      DNA      BCT      26-SEP-2000
DEFINITION      Mycobacterium avium subsp. paratuberculosis insertion sequence
15900, Locus 7.
ACCESSION      AJ250020
VERSION      AJ250020.1      GI:8919146
KEYWORDS      2599 gene; 2600 gene; insertion element; insertion sequence IS900;
SOURCE
ORGANISM      p43 gene; transposase.
Mycobacterium avium subsp. paratuberculosis.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
Mycobacterium; Mycobacterium avium complex (MAC).
1 (bases 1 to 2923)
Bull, T.J., Hermon-Taylor, J., Pavlik, I., El-Zaatar, F. and Tizard, M.
Characterization of IS900 loci in mycobacterium avium subsp.
paratuberculosis and development of multiplex PCR typing
Microbiology 146 (Pt 9), 2185-2197 (2000)
JOURNAL
MEDLINE      20431891
REFERENCE      Bull, T.J.
AUTHORS      Direct Submission
TITLE      Submitted (29-NOV-1999) Bull T.J., Surgery, St.George's Hospital
JOURNAL      Medical School, London, SW17 0RE, UNITED KINGDOM
FEATURES
SOURCE      Location/Qualifiers
          1..2923

```

gene	CDS
gene	<pre> /organism="Mycobacterium avium subsp. paratuberculosis" /sub.species="paratuberculosis" /db_xref="taxon:1770" complement(11..466) /gene="2600" complement(11..466) </pre>
CDS	<pre> complement(11..466) /gene="2600" complement(11..466) </pre>
source	<pre> /organism="Mycobacterium avium subsp. paratuberculosis" /insertion_seg="IS900" /sub.species="paratuberculosis" /db_xref="taxon:1770" 686..1906 /gene="p43" 686..1906 /gene="p43" /codon_start=1 /transl_table=1 /product="transposase" /protein_id="CAB96065.1" /db_xref="GI:8919148" /translation="MTTEVVVAQPFVAGVADGKADHYCWVINDAQRLISQVANDP AALEELIAVTTTLADGGEVFWAIDLNAGGAALLALIAACORLLYIPGRVHHAGSS YRGESKTDKDAAIADQARRRHDLQPLRACDDIAVELRLITSRSDLVADRTAIPET NAPAAIGLAAHAFEDIFPNKSRALILITGOTDALSAGAAVAAFLERKRRANDN VAAATLQANAASHIFVPGQOLAATVVARLAKFYMALDTEIGDIDMTEEFRRRAA ELIISMPGEVYILGAELFATGSGMAAFASADRLAGVAGLAPVPRDSGRISGNLRPR RYDRRLRACVYLSALVSIKRTDPSSRTIYDRKRTGKRHTQAVLALARRRLVLMALR DHAVYHPTTTTAA" complement(1963..2394) /gene="2599" complement(1963..2394) /gene="2599" /function="unknown" /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="CAB96066.1" /db_xref="GI:8919149" /translation="VSRNRLFVINGALAAVAAVCTVNGTIIILQKNIAVYAGHYHEA RDVCTGTCGSGSEGVADTLEVAERPARADDMKTEYLKSYNDIYVGPDGNPCST RVEPLSGYSHIRAVFLGPGTFPSGSGAGTTPGPGTK" </pre>
BASE COUNT	473 a 1005 c 1000 g 445 t
ORIGIN	
Query Match	80.0%; Score 20; DB 1; Length 2923;
Best Local Similarity	100.0%; Pred. No. 1.4;
Matches	20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	4 cagcaggtgcatgcgcagca 23
Db	125 CAGCAGGTGTCATGCCACCA 144
RESULT 16	
LOCUS	PSEHRPIA/c 3603 bp DNA BCF 21-OCT-1993
DEFINITION	Pseudomonas syringae hrpII genes, complete cds.
ACCESSION	L11582
VERSION	L11582.1 GI:151267
KEYWORDS	hrpII gene; protein translocation.
SOURCE	Pseudomonas syringae (strain 611) DNA.
ORGANISM	Pseudomonas syringae

REFERENCE	TITLE	JOURNAL	FEATURES
1 (bases 1 to 3603)	Characterization of the <i>Pseudomonas syringae</i> pv. <i>syringae</i> 61 hrpJ and hrpI genes: homology of HrpI to a superfamily of proteins associated with protein translocation	Plant Microb Interact. 6, 515-520 (1993)	Location/Qualifiers
1..3603			
/organism="Pseudomonas syringae"			
/db_xref="taxon:317"			
1330..3417			
/gene="hrpI"			
1330..3417			
/gene="hrpI"			
/note="hypersensitive response and pathogenicity genes, membrane-spanning protein; hrpI superfamily proteins are associated with protein translocation; putative"			
/codon_start=1			
/transl_table=11			
/product="HrpI"			
/protein_id="AAA03015.1"			
/db_xref="GI:151268"			
/translation="MNRYINFLNMVYALSAMRSELVCAFEVIAIVFMNIPPLPGILIDIVLVANICISCLITLMAHILPRPLAFSEFPVLLITTFMERLALSVTTRILLINODKSHIVAEFGVGVGNLAVGLVLEITLVTVNPLVITFSGSRVAVENGARPLDAMPKSGSIDSRANLITVHEARKRRAFLNKESOLFAGMDCAKRVNGDAIASLIVATNMGISIVGLYHNMAGDALQITVTLTGEDLAIQIPALLISTSMITVRPNEAGVEANICRELAQITISPKAMITASVAMLDLPAALIPMPGTVFITAIICGAGLILQRAKPKPADEORIAAVALPAMNKKEDLRTFSPSPROVPLQHPQDSNOIALVSEIIRKRNRLVVOYDGLTPSLPIIEHVDIADPERFEFTYDPMLEKATPTQSHVAEARQLGENLPALIPGNIDROEDQWMLPAEQSGLINPVSSTLIIEMERALOSCAQPIGLQETRAIISLESSQPEELAEQMOEVLITFRSAVLORLASCVCPLRAIRVIAETLIRCOHERTNVLTQYVRIAKSIOYHVGCGAGCQWIVITPSEBGLRGLIKQOTETTFALSNFTSOMLVQOHLIAFSPKAPRQAVLVLAQDLRSLPLRTLLREEFYHVFVLSFAETISNAKVKVMGRFDLEDDLEPLDNEHA"			
BASE COUNT	772 a 1093 c 1035 g 703 t		
ORIGIN			
Query Match	76.0%; Score 19; DB 1; Length 3603;		
Best Local Similarity	100.0%; Pred. No. 4.6;		
Matches	19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	6 gcaggtgcatgcgcagcat 24		
Db	1522 GCAGGTGCATGCAGCAT 1504		
RESULT 17			
AE007050/c			
LOCUS	AE007050 14846 bp DNA BCT 27-APR-2001		
DEFINITION	Mycobacterium tuberculosis CDC1551, section 136 of 280 of the complete genome.		
ACCESSION	AE007050 AE000516		
VERSION	AE007050.1 GI:13881590		
KEYWORDS			
SOURCE			
ORGANISM	Mycobacterium tuberculosis CDC1551.		
REFERENCE	Mycobacterium tuberculosis CDC1551.		
AUTHORS	Bacterioides; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterinae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.		
	1 (bases 1 to 14846)		
	Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J.F., Deboy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.		
TITLE	Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains		
JOURNAL	Unpublished		

REFERENCE 2 (bases 1 to 14846)
AUTHORS Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwin, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, M.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
TITLE Submitted
JOURNAL Direct Submission
Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES
SOURCE 1.14846
/organism="Mycobacterium tuberculosis CDC1551"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="clinical strain"
145..1302
/gene="MT1935"
145..1302
/gene="MT1935"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46208.1"
/db_xref="GI:13881591"
/translation="MRCENLDIYVIGISITPTTIGWVLAEGHAGDAILDRENELEHSG
RNAQAIHTAEQLAEVLAEHVAAGDHRLVIGVWNAESAQAALLVSLTGACED
NVAVRRIRAEIETLAQIAPIVGYEOLAVCELESATVYVMDTHDKTOIAVKNVR
GISGLTSMTEGMDAMPAGVYVVSDESEVSFEQSOLERVLPVFAQTMAQVVA
RGALAAOSREFPDADOLVADSOSPIVAPRSRHYGAALAAAVTTVASISLVA
GIQLAPHNDTGTAKHCAHKKPRIRAKAVAAPVPPVTPVPAPARPAQHEPPAR
VTSGEALTEPPEEQPVASAPQODRDSQDITVLEHTIGAVGDSAPPAE"
complement(1272..2237)
/gene="MT1936"
complement(1272..2237)
/gene="MT1936"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46209.1"
/db_xref="GI:13881592"
/translation="MPRTWITAGNLYRPLSNQMPIMDYOCRRHREPSCLQPKAJTA
RSGVRSARMLRCADEKROMANIVPADAHVNEPSTIMOPDAYPRVRCGLDPLSR
WQMLKWFALAIPIVIEFLHVAVVVIVAFALITGRIPRTLPFDNVVMMWRK
VAFYALSAIGDRYPPPSIQKAEYPADEVDYPERLSGLVLIKWLLAIPHLILA
VLSGMRVFLIDPHDRVGIIMPSLLVILLVAVALLETGRYPGILYNLIVGNRMA
LRVAYVTIMRDEYPLRLDMGPREQVQPATASDYSAAGAESP"
complement(2573..2929)
/gene="MT1937"
complement(2573..2929)
/gene="MT1937"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46210.1"
/db_xref="GI:13881593"
/translation="MPRTNNDAMDLATSVGATATVAAAAVATRADNPLIDDPV
LVRAVGIDPFTWAGNITKATVDVDDPGTWGIQRILADLARTRYFDAFFRDATSAGI
ROAVILASGLDARAYR"
complement(2988..3599)
/gene="MT1940"
complement(2988..3599)
/gene="MT1940"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46211.1"
/db_xref="GI:13881594"
/translation="MAHKTRREGGRAGRSSEYSRGVSADAVWTLADSGDELVLRTGVYGR
AARLGHRLTIAMTRMOALVWMSCTDPVAGELVAEDVSFEVMRGEQVKGILSEPEKALV

gene
CDS
3653..4060
/gene="MT1941"
3653..4060
/gene="MT1941"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK46212.1"
/db_xref="GI:13881595"
/translation="MIRELVTTAAITGAIGAPVAGADPQRYDGVPGMYDASLGA
PCSSMERFTFGKPGSGAACHPPPPNPPPEATGYWVSYPLGVQGVGAPCPKPPA
AAQSPDGLPMLCLGRNQPGWTFGAGFPPEP"
4077..4388
/gene="MT1942"
4077..4388
/gene="MT1942"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46213.1"
/db_xref="GI:13881596"
/translation="MIMCEGRPTESPIRPMILRFVLTSDRAGAWYTGAFPEAPYLA
VLSWPIITVAVLWIIIGLGLWGLGIGIAVAVGLARLSGAEIPPAVYRTLVDRSAN
E"
4398..4616
/gene="MT1943"
4398..4616
/gene="MT1943"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK46214.1"
/db_xref="GI:13881597"
/translation="MSFNPDAVADVADIAANAVKASDVENACHIIIRGDIAGASG
IVKDSIDIATHAVDRITKEVFTGKTDEG"
complement(4651..5781)
/gene="MT1944"
complement(4651..5781)
/gene="MT1944"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK46215.1"
/db_xref="GI:13881598"
/translation="MPTAICDELGIEFFIPATFHCADVVAVSKAGGFGVLAVGFTFP
EQLIELNIDENHIDGHPGVGIVIPKRYGMSQSLADELAKTLKLSMPOEHLIDAR
KIADHGVPEVDEDESLDGLVTEATPTGVOVAAALKHPKRTAVANAAGTPPADMTKH
IHDSGRVVAALCGSPSQAARKHAAGVNDITIIAOCGEGGHCGEVSYILMPOVYEVAP
VPIVLAAGIGSGQOIIAALALGTGAGTSGQMLVBEANATVAQOAAVYVATSRDVR
SRSTYTGPARMLRNDWTEAMEQPEPSPLPQIYNVSGMAVATATKTPRETYDAFN
PVGOVGOFTKEVTATVIERWQVEYDEATARDALNAASV"
5854..6285
/gene="MT1945"
5854..6285
/gene="MT1945"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46216.1"
/db_xref="GI:13881599"
/translation="MANSFGSRVSATPAKTPRSTVTFRRAMPASNNGCPTEOKY
WAMVCTAVGYHRDDPAELILRNEGILAAVOTGLRLPATETSSQGCGRGSRQPPR
GDHSVPRDRSPNRDRADILDPDSIGPAMDLDPGPRR"
6433..7587
/gene="MT1946"
6433..7587

```

/gene="MT1946"
/Note="similar to GB:J03362 SP:P14940 PID:141900;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="zinc-binding dehydrogenase"
/protein_id="AAK46217.1"
/db_xref="GI:13881600"
/translation="MRAVVDGAGSVRVNTPDPALPGDGVVAATVAAIGCSDLIH
TEGEVFTPEPAVAGHGAAGTIVEAGPQVRIAGVDLWSSVACGVCGETHPDM
CEGSPMIGPAGVAGCAQADLLVPAADPOVLKIPGEGITTEGALLTNNATGMAAR
ADISFGSAVAVIGAGVGLCALRSFATVAVAVRKGRLORATKATATPSPA
AETTLAATRGAGSDVDAVGTDSMSALNAVRGCGVSIVGYHDQPPPLATLCL
LRSTLRMTMAVORTPELPILOSGLVDGJFTTTLLPDEAKGTATARSSEE
complement(7447, .8358)
/gene="MT1947"
/complement(7447, .8358)
/Note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK46218.1"
/db_xref="GI:13881601"
/translation="MTPEYGLSRDSDDDHMDIVSNVGYTALLVAGMRAHTTGPPIV
ODEYAKHITASADPYLEGLANPRISDGTAPRLYGVOPRDEFFNCADEAGIKV
AVTAAIGDCRAVRLDMOPGTVEIDVPEVLEKARVLSERGAVRRAHVAVPADIR
TDMPTPLAAGFPDRPSANSVEGLPYLTGDQVYLFARIDELCAVPSVVALAGS
RUDHFOALAEHTAHGVMGSDVNSFALTYDDKTDVPEMLVEHGMADVPRSTLEQV
GYGLTPDVVKKIDSEFMSQYITAVRA"
/complement(8363, .8848)
/gene="MT1948"
/complement(8363, .8848)
CDS
Query Match 76.0%; Score 19; DB 1; Length 14846;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 cgcagcaggtgcatcgcca 20
|||||
Db 12877 CGCAGCAGGTGTCATCGCCA 12859

RESULT 18
MTCY180/c 44201 bp DNA BCT 03-AUG-2001
LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 85/162.
DEFINITION 297193 AL123456
VERSION 297193.1 GI:3261816
KEYWORDS
SOURCE Mycobacterium tuberculosis H37Rv.
ORGANISM Mycobacterium tuberculosis H37Rv.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 44201)
REFERENCE
AUTHORS Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
Harris,D., Gordon,S.V., Eigmeier,K., Gas,S., Barry III,C.E.,
Tekala,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
Connor,R., Davies,R., Devlin,K., Fellwell,T., Gentles,S.,
Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
Squares,S., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S., and
Barrell,B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
JOURNAL Nature, 393 (6685), 537-544 (1998)
MEDLINE 98295987
REFERENCE 2 (bases 1 to 44201)
AUTHORS Parkhill,J.
TITLE Direct Submission

```

JOURNAL

COMMENT

Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhillsanger.ac.uk On Jun 27, 1998 this sequence version replaced gi:2252942.

Notes:

Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.

(URL: <http://www.sanger.ac.uk/Projects/M.tuberculosis/>) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.

gene prediction was based on a Hidden Markov Model of TB genes implemented in TParse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.

CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtc, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES

```

source
Location/Qualifiers
1..44201
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
<1..42142
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="Y180"
23..460
/gene="Rv1873"
23..460
/gene="Rv1873"
23..460
/Note="Rv1873"
/Note="Rv1873"
score is 0.909
/transl_table=11
/product="hypothetical protein Rv1873"
/protein_id="CAB10052.1"
/db_xref="GI:2225987"
/db_xref="SPTREMBL:007756"
/translation="MKSADPFDLKRVVAQAPVRSVVELRAGRKGMHMFVPO
LKGSSPLAVRYSISLEQAQVLAQDLGLPRHETGTVNOVGSRISBEIGPPDD
LKSSMTLEFARVANDQFVALAKYGGGDRVALLAV"
517..520
/Note="possible RBS, GCAG, for Rv1874"
533..1219
/gene="Rv1874"
533..1219
/gene="Rv1874"
/Note="Rv1874, (MTCY180.44c), len: 228. Unknown, TParse
score is 0.928"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1874"
/protein_id="CAB10067.1"
/db_xref="GI:2225986"
/db_xref="SPTREMBL:007755"
/translation="MLMRPDDDKCARQAVADALLGICVAGISINVRDSTVPSL
MTLITLYPPAAVAVLMTQCYEYVAARLDAQDCDEGATVTSVPLTFPSLVE
SGSRTGLANIALLRPDDGDAITLTKMQCHTOVAIAQATFGYTMWVVRALPTE
APGIAIGVELFVFAATTDIKAFGAADNDLRNRISRNVASTSAFGANQNDIVPTS
RVYRTPFKD"
1218..1221
/Note="possible RBS, GAGG, for Rv1875"
1230..1673
/gene="Rv1875"
1230..1673
/gene="Rv1875"
/Note="Rv1875, (MTCY180.43c), len: 147. Some similarity

```

to295584|MTCI65.22 (147 aa) opt: 178 z-score: 240.3 E(): 7.4e-06; 26.9% identity in 130 aa overlap. Tbpase score is 0.908"

/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1875"
/protein_id="CAB10051.1"
/db_xref="GI:2225985"
/translation="MTTLEAAALAAERGLAVSTVRADGTVOASLVNGLLPHVPS
GPGSLGTTGKVKGLRPARPOLAATFRNGQMOMAVEGKALVDPDPRPWLVDGER
LRLLRPFPAAGTHDDMDYDRVMAQEQRAVVLLTPRTYSNG"

2179. .2183
/note="possible RBS, AGCAG, for Rv1876"
2189. .2668
/gene="bfrA"
2189. .2668
/gene="bfrA"
/note="Rv1876, (MTCY180.42c), bfrA, len: 159. Function:
bacterioferritin. FASTA results: BFR_MTCLE_P43315
bacterioferritin (bfr) (159 aa) opt: 958; E(): 0; 90.6%
identity in 159 aa overlap. Tbpase score is 0.913"

/codon_start=1
/transl_table=11
/product="bfrA"
/protein_id="CAB10050.1"
/db_xref="GI:2225984"
/translation="MGGDDPYRLINEQSLSELTAI NQYFLSHKMQDNGFTEIAHT
RAESFEDERHAEFTDRILLDGLPNYRISLRIGTLEKPEFADAIEDVLRNLRK
PGIWMCKRQDTSVAVLEKIVADEEHIDYLTQLEIMKLGSELYSACVSRPPT"

2742. .2746
/note="possible RBS, GCAGC, for Rv1877"
2753. .4816
/gene="Rv1877"
2753. .4816
/gene="Rv1877"
/note="Rv1877, (MTCY180.41c), len: 687. Integral membrane
protein similar to many antibiotic and drug efflux
proteins. M. tuberculosis relatives include: MTCY312.01,
MTCY98.0002c and MTCY369.27c. Contains P500217 Sugar
transport protein signature 2 (P500217). FASTA results:
056175 (557 aa) opt: 895; E(): 0; 34.7% identity in 528
aa overlap. Tbpase score is 0.916"

/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1877"
/protein_id="CAB10049.1"
/db_xref="GI:2225983"
/translation="MAGPTAPPTAPTAIRAGGPIISPVRNITFALVFGVLAATGQ
TIVPALPTIVALELSTVDQSWAMVSYLLGCVVVVAGKLGILLGRNRVLGSAVVE
VNGSVLCGLSOTMTLAISRALOGVAGCAISTATVLAALAEVPIPLRRCGVCAV
GVNATVGLGLMTDLYLSRWAEWINVPSVLAIVATVAPALRRPKPTIDYLCI
LVIAATTAALIMATSMGCTTVAAGSATVIGLIGAVAGVLEGRPAALIPRL
FESPVAACCVSLFVFGAMLGALTFVPIYLVGASATGSLTLEPMVIGLLAT
GTVLVGRTRGKIFPVAGMALMAVAFLILMSQMDMPPLQSLTILVILVAGISLSMO
VILVIVNNTSPFEDIGVATSGVTFPFGVASCATGATFGLAEVNFIDRLRGLISGAV
EPVAPSPAVILHQLPQSMAPVIVRAVESTLQVLCASVTVVGGTLLALLREVPLT
IHDDADLDGCGVRAESPDEVLELAVNRMLPNCVRLPDTATGDCGCGGLAVELMAL
RIVYQRFEEAVRLTIDGRHLPVPOVEPFDRLQVGTIARDCDILTLTPSGRQV
DSIAVLIRQMLDLHLAVAPGLRKRPDHOFEALQHVTDALVQRMDYEDLGLSRSKO
LAAT"

3122. .3199
/gene="Rv1877"
/note="P500217 Sugar transport proteins signature 2"
4871. .6223
/gene="glnA3"
4871. .6223
/gene="glnA3"
/note="Rv1878, (MTCY180.40c), len: 450. glnA3. Similar to
many glutamine synthetases. FASTA results: GlnA_BACCE
P19064 (EC 6.3.1.2) (443 aa) opt: 497; E(): 5.2e-23; 29.0%

identity in 331 aa overlap. Tbpase score is 0.925. Also
similar to C-terminus of Fluc_EMENT p38094 fluc protein.
emericalianidians (865 aa). fasta scores: opt: 227
z-score: 357.9 E(): 6.4e-13 (29.9% identity in 394 aa
overlap) (note that the downstream ORF MTCY180.39c is
similar to the N-terminus) and to several other M.
tuberculosis proteins eg. MTCY427.03c, MTCY427.01 and
MTCY190.31"

/codon_start=1
/transl_table=11
/product="glnA3"
/protein_id="CAB10048.1"
/db_xref="GI:2225982"
/translation="MTATPLAAALIAOLEAEVDVITGVVNPAGLTQAKTVPTRTN
FRANGICASPVMHFCIDQCSIAFTPAIDSVGGDRLDLSALRIIDGLAMPAGF
PRDGTVPVACSRGTLSRIEALADAGIDAVIGVEFLVDADGQRIPTLMQYGV
AGVLEHAEFVRVNAATATAGTAIDQHFIBRGANQFETSLAPQPVAAADQVLRLL
IGTRARHGLRVSLSPAPAGSISGAIDHESLHSEGLSGGTGAAGMSAGBAV
ACVLRGLPDAQGLIGSIVSGLRMRPGMAGIYACWGTENRAAVRFGGAGSVGG
NVEVKVDPSPAPYLAASAILGLAIDCKMTKAVLPSETVPTQLSDVDRRAGILRL
AADQDAIVAVLSDSKLRLCIIIDPVDVAVAVRQLEHBERYGLDPAQLADKFRMAVSV
"

6213. .6216
/gene="glnA3"
/note="possible RBS, GCAG, for Rv1879"
6226. .7362
gene

RBS
6213. .6216
/gene="glnA3"
/note="possible RBS, GCAG, for Rv1879"

Query Match 76.0%; Score 19; DB 1; Length 44201;
Best Local Similarity 100.0%; Pred. No. 3.2; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0

Oy 2 cgcagcagtcgcatcgcca 20
Db 25769 CGCAGCAGTCGTCATCGCCA 25751
|||||

RESULT 19
AC068068/c
LOCUS
DEFINITION
AC068068 37500 bp DNA HTG 28-APR-2000
Leishmania major chromosome 2 clone L5666 strain Friedlin, ***
SEQUENCING IN PROGRESS ***, 11 unordered pieces.
AC068068
VERSION
AC068068.1 GI:7658330
HTG: HTGS_PHASL.
KEYWORDS
Leishmania major.
SOURCE
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
ORGANISM
Leishmania.
1 (bases 1 to 37500)
Miyler, P.J., Sisk, E., Ruiz, J., Cosenza, P., Cruz, A. and Stuart, K.
Direct Submission
Submitted (28-APR-2000) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 12249: contig of 12249 bp in length
* 12250 12296: gap of unknown length
* 12297 27295: contig of 14999 bp in length
* 27296 27342: gap of unknown length
* 27343 27885: contig of 543 bp in length
* 27886 27932: gap of unknown length
* 27933 28468: contig of 536 bp in length
* 28469 28515: gap of unknown length
* 28516 28577: contig of 142 bp in length
* 28578 28703: gap of unknown length
* 28704 29418: contig of 715 bp in length

```

*      29419      29464: gap of unknown length
*      29465      30625: contig of 1161 bp in length
*      30626      30671: gap of unknown length
*      30672      31824: contig of 1153 bp in length
*      31825      31870: gap of unknown length
*      31871      33509: contig of 1639 bp in length
*      33510      33555: gap of unknown length
*      33556      35178: contig of 1623 bp in length
*      35179      35224: gap of unknown length
*      35225      37500: contig of 2276 bp in length.
*
FEATURES
  source
    1. 37500
    /organism="Leishmania major"
    /strain="Friedlin"
    /db_xref="taxon:5664"
    /chromosome="2"
    /clone="L5666"
BASE COUNT      6965 a 11364 c 11233 g 7435 t 503 others
ORIGIN

```

```

Query Match      72.0%; Score 18; DB 2; Length 37500;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 ggcgcagcagtgcatcgc 18
Db      5534 GCGCAGCAGTGTCATCGC 5517

```

```

RESULT 20
AC010079/c      38000 bp      DNA      HTG      11-SEP-1999
LOCUS
DEFINITION      Leishmania major chromosome 2 clone L4116 strain Friedlin, ***
ACCESSION      AC010079.1 GI:5870277
VERSION
KEYWORDS      HTG; HTGS_PHASE2.
SOURCE      Leishmania major.
ORGANISM      Leishmania major
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

```

```

REFERENCE      1 (bases 1 to 38000)
AUTHORS      Myler,P.J., Sisk,E., Ruiz,J., Cosenza,P., Cruz,A. and Stuart,K.
TITLE      Direct Submission
JOURNAL      Submitted (11-SEP-1999) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
COMMENT
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 3 contigs. Gaps between the contigs
  * are represented as runs of N. The order of the pieces
  * is believed to be correct as given, however the sizes
  * of the gaps between them are based on estimates that have
  * provided by the submitor.
  * This sequence will be replaced
  * by the finished sequence as soon as it is available and
  * the accession number will be preserved.

```

```

*      1079      1201: gap of 1078 bp in length
*      1202      22879: contig of 21678 bp in length
*      22880      23001: gap of unknown length
*      23002      38000: contig of 14999 bp in length.
*
FEATURES
  source
    1. 38000
    /organism="Leishmania major"
    /strain="Friedlin"
    /db_xref="taxon:5664"
    /chromosome="2"
    /clone="L4116"
BASE COUNT      6861 a 11830 c 11904 g 7132 t 273 others
ORIGIN

```

Query Match 72.0%; Score 18; DB 2; Length 38000;

```

Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 ggcgcagcagtgcatcgc 18
Db      16164 GCGCAGCAGTGTCATCGC 16147

```

```

RESULT 21
AC087836/c      42000 bp      DNA      HTG      28-FEB-2001
LOCUS
DEFINITION      Leishmania major chromosome 2 clone B4162 strain Friedlin, ***
ACCESSION      AC087836.2 GI:13162520
VERSION
KEYWORDS      HTG; HTGS_PHASE1.
SOURCE      Leishmania major.
ORGANISM      Leishmania major
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

```

```

REFERENCE      1 (bases 1 to 42000)
AUTHORS      Myler,P.J., Sisk,E., Ruiz,J., Cosenza,P., Cruz,A., Stuart,K.,
Dobson,D.E., Beverley,S.M. and Scholtes,L.
TITLE      Direct Submission
JOURNAL      Submitted (30-JAN-2001) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
COMMENT
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 4 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.

```

```

*      1      8056: contig of 8056 bp in length
*      8057      8289: gap of unknown length
*      8290      18958: contig of 10669 bp in length
*      18959      19191: gap of unknown length
*      19192      40819: contig of 21628 bp in length
*      40820      41051: gap of unknown length
*      41052      42000: contig of 949 bp in length.
*
FEATURES
  source
    1. 42000
    /organism="Leishmania major"
    /strain="Friedlin"
    /db_xref="taxon:5664"
    /chromosome="2"
    /clone="B4162"
BASE COUNT      7450 a 12790 c 13550 g 7510 t 700 others
ORIGIN

```

```

Query Match      72.0%; Score 18; DB 2; Length 42000;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 ggcgcagcagtgcatcgc 18
Db      34508 GCGCAGCAGTGTCATCGC 34491

```

```

RESULT 22
AE004192
LOCUS      AE004192      12891 bp      DNA      BCT      31-JUL-2000
DEFINITION      Vibrio cholerae chromosome I, section 100 of 251 of the complete
chromosome.
ACCESSION      AE004192 AE003852
VERSION
KEYWORDS      AE004192.1 GI:9655581
SOURCE      Vibrio cholerae.
ORGANISM      Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

```

REFERENCE
AUTHORS
1 (bases 1 to 12891)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D., Nierman, W.C. and White, O.
DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*
Nature 406 (6795), 477-483 (2000)

TITLE
JOURNAL
MEDLINE
2 (bases 1 to 12891)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D., Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R., Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
Direct Submission
Submitted (14-JUN-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

FEATURES
Source
1. 12891
/organism="Vibrio cholerae"
/strain="N16961"
/serotype="O1"
/db_xref="taxon:666"
/chromosome="1"
/note="biotype: EL 'Tor'"
complement(84..1370)
/gene="VC1111"
complement(84..1370)
/note="VC1111"
/note="similar to GB:J04423 SP:P12995 PID:457106
PID:490217 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="adenosylmethionine-8-amino-7-oxononanoate
aminotransferase"
/protein_id="AAF94270.1"
/db_xref="GI:9655582"
/translation="MMMDLAFDRCHIMHPYSTLPLCYGYPVSRADGYMTLLEDSRL
VGGSSWMAAIGHYNHKPLNOAAHOIEMSHIMGGITLHPATELCOKLKLAENNL
QQVFLADSGSAVEVSLKMAIDYHMSKOPRAKPTLHRGYCHDFEAMSTDPNSM
HSUYKGFLEPHFANSPGEGFDWADERDIAFRKLELHHQIAVILLEPLVQACG
MRITYPEFLRQVRAIDCEGVLLIDELATGFGKLFACCHAGIQDPILVGRALT
GGYMTLSATLTQQVADIVCAGAGCFMNGPTFMGNPLACAVASASLSIEGDWQOO
VANIEAFEAELPKLNDSPRVKOTRWLAGIVGVEHLDPVMEVIALFEVHGVMTRPF
GRUYIMPPYISQPEHLKLVDAAMAALNQPCFKQ"
1497..2549
/gene="VC1112"
1497..2549
/note="VC1112"
/note="similar to GB:J04423 SP:P12996 PID:145425
PID:490219 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="biotin synthase"
/protein_id="AAF94271.1"
/db_xref="GI:9655583"
/translation="MEVRNMTVAEYKALLDKPPMDLFEAQVHRLHHPHHVQVST
LISITKACPEDCCKYCPOSAHYRTDVKRLEMEVERLDAOKAKNSGSRIPCMGAM
KNKERDMLLEKMRQVKMGLETCMGLTPDOAOOLAAGLDIYNNLDTSEF
YGNITTRTYQDRDLTLISHVRDGMKICSGGITTGMGESTNDRAGLIVELANLPTPES
VPIIMLVKVGKPLEVDVPEFVNLVARIIMPKSAVRLSAGREKKNEDMOALC
FMAGANSIPYCKKLTTPNPDAEDSDMLPKKIGINREQVAKPKDEITENLLDRVER
VAARPASDLEFYDAAL"
2533..3687
/gene="VC1113"
2533..3687

gene
CDS
/gene="VC1113"
PID:490225 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="8-amino-7-oxononanoate synthase"
/protein_id="AAF94272.1"
/db_xref="GI:9655584"
/translation="MTLRKSRIDEALNREHRSRLTRRLQILDRSAGCELIOODRTYL
NFSSNDYGLDANDAEVLQAWOTGLARFGSGSASPMYGFSSAHLEHLECHMGE
RAVLFFSGFSANQALFTLLEKEDLLDDKLNHASLMEAGMLSPATMKRKHNTDHL
ROLHEOSNSLVNTEGVFSMDGCAPLQIATLVQRAMLVWDANHGIGVGEBSAGS
COAGCHPEILVTEFGKAPGLAGAAVLCDHAGDYLTQFARIHVSSTAMPQAQHALT
HAARQOSQSWRREOLAEACFDQCRNIPGVVLTQPIKPMILGSSDSALSASHEL
KQOGIVNSIRPPYVYGSARLRITITTAHQALIRLSSEOLKHYMERKT"
3549..4487
/gene="VC1114"
3549..4487
/note="VC1114"
/note="similar to GB:J04423 SP:P12999 PID:145427
PID:490221 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="biotin synthesis protein Bioc"
/protein_id="AAF94273.1"
/db_xref="GI:9655585"
/translation="MGECSTTHRACROCAFTHYONGGAVPSADPPLIGAVKTCDEG
HMSKTAELCEVELKDKSAITQAFKAAHYDQAAARQVGRRLLOKMPSCGLRVL
DLGGTGYSALLREKQAOVYCADISHAMLEOKKQKCGDMSYOLADAEOLPPASAC
FDWVFSSIALQWCEDELSLSEITRVKLPBQCAFSTLIDGSLFELEQMRSDHHRH
INQFISINQVRIALQAGCSQHHLDLAITVWEAFMLMRDLKIGAHNVSGRSTGL
LSRRTAKVELAVQSFRRNOGLVPATYQVCLGVYHR"
4463..5179
/gene="VC1115"
4463..5179
/note="VC1115"
/note="similar to GB:S68059 SP:P13000 PID:490223
PID:145428 PID:460901; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase"
/protein_id="AAF94274.1"
/db_xref="GI:9655586"
/translation="MFRGYSMPRNAFIAGTDDVGKTVASKALQALATHTNATGTY
KPVASGSKTEFGYRNSDALHMKAAATVDMPEEDVNPYALVLPSPHIAKHENVTD
YALLSNKTSKKNQAEILVYVGGAGGMRVPSDSCLSTWKOERLPVILVVGKLGCL
SHALTRATADLELTVGWIANRINGTENYALIEHLGRLGTPKLGELPYMPKAK
ROELGKFIQDHLLEPDTVA"
5261..5470
/gene="VC1116"
5261..5470
/note="VC1116"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF94275.1"
/db_xref="GI:9655587"
/translation="MSRRNLTTNKGKHAHAASYOMATDLLCCHLGISSEQARTQGL
EAKRQTIQOQIVDTQHSIMLSAEK"
5697..6560
/gene="VC1117"
5697..6560
/note="VC1117"
/note="similar to GB:M58470 SP:P23894 PID:146412 GB:U00096
PID:1736470; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="heat shock protein HspX"
/protein_id="AAF94276.1"

```

/db_xref="GI:9655588"
/translation="MKRLLFLATNLAVVLVSVLNIYVAVTGMQSGLSGLLWMA
VFGGATISLMSKSMALRSVGVVDTPPNEMHLEETVRQADAGIGPTVAI
IDAPMNAFATGARDSDSLVAVSTGLANMRDEAVLAIEVSHIANGDVPTLMO
GVNFTVFELFRFANIANVASRDESGSNMAYEGVSMVEIVGEGFLASTITWYSR
HREFPDAGAAQLVQKHMALERLKMGQSHLEGSMAFGCTGKSLSLMTHTPP
LEKRIALRN"
6770..7198
/gene="VC1118"
/note="Similar to PID:1653608; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, putative"
/protein_id="AA094277.1"
/db_xref="GI:9655589"
/translation="MSINDVQVNAQFYQOLDKSQLRLTEIHPDVVEFAARIECP
DALQYFLNLXQNVHTCTPIIHQYAVNAGAFIWMHILRPPKLAKGEVDVAGVSHI
HFAEGKVTYHRDYMGMELVQLPVIGVYRAIKRRLGQ"
7195..7920
/gene="VC1119"
/locus="7195..7920"
/note="VC1119"
/note="similar to SP:P25970; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="oxidoreductase, short-chain
dehydrogenase/reductase family"
/protein_id="AA094278.1"
/db_xref="GI:9655590"
/translation="MTAVFTGATSGIGCOLADYAKGQMOYIACGRNSVLSLHTQ
YANIPLEPVDTHPGTKAKALQPCQPELMIINADCEITIDGMDVTLMAVFEN
VLSAVVIEGTQPHLSGHRVAIVGSIASBLPRAEAYGASKAAVAYLARTLODMR
PLIEVTTTFPGVATPPLTDNTPAMPMTIIVRAQELKAGIARGVSOLVPPKRFY
LRLCALPILMGRLVRLRLKA"
7931..9211
/gene="VC1120"
/locus="7931..9211"
/note="VC1120"
/note="identified by Climmer2; putative"
/codon_start=1
/transl_table=11

Query Match      68.0%; Score 17; DB 1: Length 12891;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1  ggcgcagcagtgcatcg 17
Db      1575  GCCGACGAGTGTCATCG 1591
|||||

```

RESULT 23
AK024397/c
LOCUS
DEFINITION
AK024397 4674 bp mRNA PRI 29-SEP-2000
Homo sapiens CDNA FL014335 fis, clone PLACE4000411, highly similar
to Homo sapiens mRNA: CDNA DKFZ586D0624 (from clone
DKFZ586D0624).

ACCESSION
AK024397
VERSION
AK024397.1 GI:10436775
KEYWORDS
oligo capping; fis (full insert sequence).
SOURCE
Homo sapiens placenta CDNA to mRNA, clone: PLACE4
clone: PLACE4000411.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (sites)
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shitatori,A., Sudo,H.,
Wagatsuma,M., Hosotani,T., Kaku,Y., Kodaira,H., Kondo,H.,

```

Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,  
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,  
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,  
Nakamura,Y., Nagahara,K., Masuo,Y., Nimomiya,K. and Iwayanagi,T.  
NEDO human cDNA sequencing project  
Unpublished (2000)  
2 (bases 1 to 4674)  
Isogai,T. and Otsuki,T.  
Direct Submission  
JOURNAL  
Submitted (23-AUG-2000) to the DDBJ/EMBL/Genbank databases. Takao  
Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana,  
Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@helix.co.jp,  
Tel:81-438-52-3951, Fax:81-438-52-3952)  
NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan: cDNA full insert  
sequencing. Research Association for Biotechnology: cDNA library  
construction, 5'- & 3'-end one pass sequencing and clone selection:  
Helix Research Institute (supported by Japan Key Technology Center  
etc.) and Department of Virology, Institute of Medical Science,  
University of Tokyo.  
Location/Qualifiers  
1..4674  
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/clone="PLACE4000411"  
/clone_lib="PLACE4"  
/tissue_type="Placenta"  
/note="Cloning vector: pME18FL3"  
608..1273  
/note="unnamed protein product"  
/codon_start=1  
/protein_id="BAB14908.1"  
/db_xref="GI:10436776"  
/translation="MLPSIKYLGWTLVEQPKGEELSAATIKRIYATKASGKLQYK  
LKVSPTGILINDLJNOLIEVSTYRSTCADKMKHRYATINQSHNSLCHAF  
CTKRMAQVITLYAQAFKVAEPFQVSKERKRDASQGGVIGARQDCPPPLKS  
LVACGNLIDLETKAPLSTVSAWTTMDEVPPOALSSGVVWSCVASRSVITSL  
TSG"

BASE COUNT      991 a 1342 c 1239 g 1102 t
ORIGIN

Query Match      64.0%; Score 16; DB 9: Length 4674;
Best Local Similarity 100.0%; Pred. No. 1,6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      9  ggtgcacgcagcat 24
Db      2446  GGTGCATGCCGACGAT 2431
|||||

```

RESULT 24
AE005685
LOCUS
DEFINITION
AE005685 Caulobacter crescentus section 11 of 359 of the complete genome.
VERSION
AE005685.1 GI:13421206
KEYWORDS
Caulobacter crescentus.
SOURCE
Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
Caulobacter.
1 (bases 1 to 10290)
Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E.,
Eisen,J., Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R.,
Potočka,F., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D.,
Ely,B., Debby,R.T., Dodson,R.J., Durkin,A.S., Gwin,M.L.,
Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Kouri,H., Shetty,J.,
Berry,K., Ueberback,T., Tran,K., Wolf,A., Vamathevan,J.,
Emolaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and
Fraser,C.M.
Complete genome sequence of Caulobacter crescentus
Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)

MEDLINE 21173698
 2 (bases 1 to 10290)
 REFERENCE
 AUTHORS Nieman, W.C., Feldblum, T.V., Paulsen, J.T., Nelson, K.E., Eisen, J.,
 Heiderberg, J.F., Alley, M.R.K., Ohta, N., Maddock, J.R., Potocka, I.,
 Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B.,
 Laub, M.T., Deboy, R.T., Dodson, R.J., Durkin, A.S., Ginn, M.L.,
 Hall, D.H., Kolonay, J.F., Smit, J., Craven, M., Khouri, H., Shetty, J.,
 Berry, K., Ulebrack, T., Tran, K., Wolf, A., Vamaliavan, J.,
 Ermoiaeva, M., White, O., Salzberg, S.L., Shapiro, L., Venter, J.C. and
 Fraser, C.M.
 TITLE Direct Submission
 JOURNAL Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 FEATURES
 source Location/Qualifiers
 1. 10290
 /organism="Caulobacter crescentus"
 /db_xref="taxon:69394"
 complement(74. .1363)
 /gene="CC0102"
 complement(74. .1363)
 /note="identified by match to TIGR protein family HMM
 TIGR00563"
 /codon_start=1
 /transl_table=11
 /product="rRNA methyltransferase RsmB"
 /protein_id="AAK22089.1"
 /db_xref="GI:13421207"
 /translation="MTGLNDGLPAREALTLIDALSRGGLDEASANGRELEPR
 ERAFARALAMAVLRIGPIDRALITARILOKAPPRVMI.LRLGAQAFYLEVPFAVA
 TSVELAGANKTSRPFKGLVNAVLRGLRDAPADPSLAPMLYARWVALEGEATAR
 QIAGCIATEPTDLSARGLADLAPLALFALAEVLPGGTURLRKGDAWAMPGEEDGTW
 WQDAAAAPARLLIDVOPGASVYDLCAPGSKTLOLAACGAVALDRAARLRKLSL
 NLTRGLSAETVAADAAMWDITRTFDLILDAEPSATCTPRRHVDVILMARPGDASL
 ANVORILDAADRLKPGARIVCVCSLEPEGCAOYEAFLARODVALEPITRAADAP
 SAGEPESLTPRGTLRLPHHRDDGDFEAFARRKL"
 1426. 1704
 /gene="CC0103"
 1426. 1704
 /gene="CC0103"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK22090.1"
 /db_xref="GI:13421208"
 /translation="WVIVGGDDRSITRQCAAMTDTPPTDAPPYCAAPKCTLTP
 AARRALEFAARRAEALSLPEEGGRDPEPTRFQDWKGVAVDF"
 1738. 3489
 /gene="CC0104"
 1738. 3489
 /gene="CC0104"
 /note="identified by match to TIGR protein family HMM
 TIGR00066"
 /codon_start=1
 /transl_table=11
 /product="gamma-glutamyltransferase"
 /protein_id="AAK22091.1"
 /db_xref="GI:13421209"
 /translation="MPAPPRLAMRNDLLEALRLFAPLAALSALLLVPSYAKP
 RSLAAAVRALLEGAVASPDQYALAAQAEVLAKGAVADVAAVATGFLAVTPEGNLQ
 GGGFTVYDGRKPYELDYRETPAARADMYLGPDELVEGLSLVGLAAGVAPCTVRG
 LAMAHORFGKLPWGOVLAPATRYARDGFVYDAHLAGVLADEPPFEATINFRYFAAH
 KAGTEKPEBELATLERIALVNGDRAPECKTADLVAOMARGPVKGLTTKADLSYKA
 AMRTPLVAKRGYDVTTPAPSSGGLALMQLLMKONLAPMGVALNDPOYVHLAE
 IERKRVADRAEYLDGDFVKVSNLIDVAYARRAKEDYPAKPSATKAVOPIGLKHE
 TTHYSTVDQMGNAVNTYVNLNGSFGSVVVEGAGFTLNNEMDVESAKGALNMGCVG
 GAANALAPKRPILSMTPTLITKDRVAVAVIGTPGSRIFEFVAVFQVLNHDGLSLK
 AAOAPRPHHOLLPONVIFYERSGPPSADLKALBARGYTWESFGDVEAIOYLGRT
 PVPEPPRARGVGLYK"
 complement(3618. .4319)
 /gene="CC0105"
 complement(3618. .4319)
 CDS

/gene="CC0105"
 /note="identified by match to protein family HMM"
 /codon_start=1
 /transl_table=11
 /product="orotidine 5'-phosphate decarboxylase"
 /protein_id="AAK22092.1"
 /db_xref="GI:13421210"
 /translation="MTADPRLVPLDLPPTVDARAVRGLDASFTYIGIELLASG
 MGLAHDKSGSKIPLMDKLHDIGATVERSAVILATSGCDLTVHAEPOVQAANAKR
 GDSKLILAVTVTLTDLADLVEMGYAFARPLVERRVOALACGVDSIVSPHEAL
 ARELANDAGPREFLIYTPGVPRFWSAKKNQAAATPADALRAGATHLVCGSPITAND
 PREALAVYEMAGT"
 4417. 5115
 /gene="CC0106"
 4417. 5115
 /gene="CC0106"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAK22093.1"
 /db_xref="GI:13421211"
 /translation="MTEANLVNFEVVALPDLPIGNVPLFAAATLGAAGAAAGRWAV
 YIGLFMAFLIFFTGVLGFEFGISMPAFRIAGIILFIILGLMDVRDFTTFADA
 AEGTEDSPRAYAKORPERLIVPAMPLITIGPSITVITYAAEAKDFGAAGIGIVA
 IANASLVTLTTEWASPLVSKVLRIGMSIVNVGLIICALAVQFVLGVADATRL
 LKNDAPKAYAESHK"
 5199. 6428
 /gene="CC0107"
 5199. 6428
 /gene="CC0107"
 /note="identified by match to PFAM protein family HMM
 PF01360"
 /codon_start=1
 /transl_table=11
 /product="2-octaprenyl-6-methoxyphenol hydroxylase"
 /protein_id="AAK22094.1"
 /db_xref="GI:13421212"
 /translation="MTTAGMGAGSTLALASGLRPLLVDPAPPEAOALPTFGRS
 SAIVSFPROWRAIGAGOLEPHAKRLEQILVYDGRTPGAARKPLAPLRDPSSTIS
 GRIEEPLGYLVENKROIKAALSKAVIDKNIPVLAIPMAKALSVEPKALTLSDGNL
 TAPILVASGSGSVLRRAAGIGDIGSGSVYCVLRLEPRPHEGVAHEYFLPSGPA
 ILPLTDNRASLVWTESTARGELRQAGEFHALMRFGDGLGAVSIEGPIFYVPS
 ISLARWAPRALIGDAAGVHPILAGGINTLGIKDAALAEVLVEAVRHGEDIGAE
 VLERYARWRREDVTNALAFDAFVVRITSNDNPLRLAKAGLAVNVNIAPARRFHME
 AGGVGDPLKLRGQAL"
 complement(6589. .6786)
 /gene="CC0108"
 complement(6589. .6786)
 /gene="CC0108"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAK22095.1"
 /db_xref="GI:13421213"
 /translation="MTESAPRLIVDPRLLEVIVCPYTRGPLEYDRAKKGELISRAKL
 AYPTRDGVPIMLPEARELEEG"
 complement(6796. .7473)
 /gene="CC0109"
 complement(6796. .7473)
 /gene="CC0109"
 /note="identified by match to PFAM protein family HMM
 PF02190"
 /codon_start=1
 /transl_table=11
 /product="ATP-dependent protease La domain protein"
 /protein_id="AAK22096.1"
 /db_xref="GI:13421214"
 /translation="MPVATRKIDDLPLVTPVPRIDGVLISGOLPLNIFPRYINML
 DDAMSGERMIGMTPTPLPGKGDQRPALAVGACAGATSEIDRATSLRITLITGVC
 RFRTEGELPVRTPYROVRADFPAYQADREDAAGRTASEIDRATSLRITLITGVC
 IDMSDAEAPSDALINSILAMALPDPMEKQALLENETIFERRATLTALLEIDAAGGD

```

gene      complement(7489. .8388)
CDS       /gene="CC0110"
          /complement(7489. .8388)
          /gene="CC0110"
          /note="identified by match to protein family HMM"
          /codon_start=1
          /transl_table=11
          /product="thioredoxin"
          /protein_id="AAK22097.1"
          /db_xref="GI:13421215"
          /translation="MSLIGEKAPSSKSDHTKDGTDASPMADVLEAKTDPVIVDEMA
          TWGPGCRLPTTEKVVTAAGAKVKKIVDKMPFRAGQIRVOSTITVAVDGRIV
          DAFQALPESQVRAFEKLAGAAGALDEITANGKESLEIGDGAQAQAFADALID
          PTKSAKAGMARVTLANGDEPGAEEVAVMAADAKDPDLDAARALAIADAPSETVA
          FEORLADPDDEARLELAKNACVAGRLDADAADHILTIARDPTWMDARKOKLTVF
          EAAGPTEVAKGRRKLSILFS"
          /complement(8498. .9004)
          /gene="CC0111"
          /complement(8498. .9004)
          /gene="VC0111"
          /note="similar to GP:6899999; identified by sequence
          similarity; putative"
          /codon_start=1
          /transl_table=11
          /product="DNA-binding protein, putative"
          /protein_id="AAK22098.1"
          /db_xref="GI:13421216"
          /translation="MKTRADLFAFPFAHGVNHTLDPVFRVEGLEIKAMPGGHT
          KNLPLKAKGOLMISALGETTIDLKLIHVLISGRLESPQEMLETTLVPGSVTA
          FGLIDTKRFRVFLDKALADSDPVNPHLPKNDATTAVSQAGLRFLAALGVPEMLVD
          PAAMEVVG"
          /complement(9081. .9827)
          /gene="CC0112"

Query Match      64.0%; Score 16; DB 1; Length 10290;
Best Local Similarity 100.0%; Pred. No. 1; Se=02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 ggcgacagatgcac 16
        |||
Db      1112 GCGCAGCAGGTCATC 1127

RESULT 25
LOCUS   MTV031 13350 bp DNA BCT 03-AUG-2001
DEFINITION
Mycobacterium tuberculosis H37Rv complete genome; segment 6/162.
ACCESSION AL021926 AL123456
VERSION   AL021926.1 GI:3261520
KEYWORDS
SOURCE    Mycobacterium tuberculosis H37Rv.
           Mycobacterium tuberculosis H37Rv.
           Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
           Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
           Mycobacterium tuberculosis complex.
REFERENCE 1 (bases 1 to 13350)
AUTHORS   Cole, S.T., Brosch, R., Parkhill, J., Garfield, T., Churcher, C.,
           Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,
           Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
           Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
           Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
           Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
           Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skellern, S.,
           Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S., and
           Barrett, B.G.
TITLE     Deciphering the biology of Mycobacterium tuberculosis from the
           complete genome sequence
JOURNAL   Nature, 393 (6685), 537-544 (1998)
MEDLINE   98295987
REFERENCE 2 (bases 1 to 13350)
AUTHORS   Parkhill, J.
TITLE     Direct Submission

```

```

JOURNAL
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:2905771.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(ORF, http://www.sanger.ac.uk/Projects/M.tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in TBparse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(ata, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.
FEATURES
Location/Qualifiers
source      1..13350
            /organism="Mycobacterium tuberculosis H37Rv"
            /strain="H37Rv"
            /db_xref="taxon:83332"
            <1..>13350
            /note="fragment designated v031: does not represent a
            physical clone"
            complement(243. .452)
            /gene="RV0108c"
            complement(243. .452)
            /gene="RV0108c"
            /note="RV0108c"
            /note="RV0108c" (MTV031.02c), len: 69. Unknown. TBparse
            score is 0.890"
            /codon_start=1
            /transl_table=11
            /product="hypothetical protein RV0108c"
            /protein_id="CAAI17302.1"
            /db_xref="GI:2909573"
            /db_xref="SPTRMBL:053630"
            /translation="MVPFTLHSGDPTTDVWGGGQRIYVLESTVGDSCVYLELSRV
            NHOVLVEKSPPAHYHGRAHRL"
            720..723
            /note="possible RBS GAGG for RV0108c"
            730..2220
            /gene="PE_PGRS"
            730..2220
            /gene="PE_PGRS"
            /note="PE_PGRS"
            /gene="RV0109" (MTV031.03c), len: 496. Member of M.
            tuberculosis PE-family, Gly-, Ala-rich PGRS subfamily.
            FASTA scores: sp|Q50615|YODP_MYCTU_HYPOTHETICAL
            GLYCINE-RICH 40.8 KD (498 aa) Opt: 1772 z-score: 1133.2
            E(): 0; 57.3% identity in 513 aa overlap. TBparse score is
            0.884"
            /codon_start=1
            /transl_table=11
            /product="PE_PGRS"
            /protein_id="CAAI17303.1"
            /db_xref="GI:2909574"
            /db_xref="SPTRMBL:053631"
            /translation="MSLITSPATVAATHLAGISALSTANMAAATPTLSVGA
            TVOOVLITVNAFTOALLRPITGNANCLPTNGDGGGGLLFNGGNGSGGVDOA
            GAGGAGLIGLFGSGGTGKGVGMGVGNGVNAAGCTTATGPGHAGAGAGIGTGTGDDPG
            GAGGAGNGRFGMGNGVGAAGYGAAGDGNAGNGTATGVPFGDAGAGTGLGADGPG
            LAGTGNAGFVRNGAGAGNGOLGSGAVSSAGMGNGTGLVFGNGPGLIGPGCTSAGN
            GGAGNAVGLFPGGAGAGGSGFAGTGGGNGGSGGGLTGDGCTGGAGAGADAAA
            SAGGNGNARLILNGGSGRPMFGPGGAGGSGGTTFPGATGPGS"
            2368..3117
            /gene="RV0110"

```

```

CDS
2368..3117
/feature="Rv0110"
/feature="Rv0110", (MTV031.04), len 249. Hypothetical
transmembrane protein similar to many e.g.
SPIP4493(YOGP_BACSU HYPOHETICAL 56.4 KD PROTEIN IN S
(507 aa) FASTA scores: opt: 320 z-score: 406.2 E(1): 4e-15;
32.4% identity in 210 aa overlap Tblast score is 0.915"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0110"
/protein_id="CAAI7304.1"
/db_xref="GI:2909575"
/db_xref="SPTREMBL:O53632"
/translation="MRGPGVHOCACVREGARVROPTPRGROSRATPVYTLI
SLNLFVYMOYVWGLEROLALMPAPVAGQYRLVTSALHGHLLNMLLYV
GPLEMNLGRLRGALYAVSALSGSVLYLAPLNTATGASGAVGLGALFMVVR
LHLDVRVVALIVINLAFPLAPRISOGVGLVTCALVAAIYVAPRERNLIGAT
VTITVLAFFVVLIGRTVDLLALEGRLNL"
3298..5355
/feature="Rv0111"
3298..5355
/feature="Rv0111", (MTV031.05), len: 685. Possible
acetyltransferase similar to bacillus subtilis
BSAB0014_194 (634 aa). FASTA scores:
gpl299117BSUB0014_194 (634 aa) opt: 520 z-score: 531.7
E(1): 4e-22; 36.4% identity in 382 aa overlap. Tblast score
is 0.937"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0111"
/protein_id="CAAI7305.1"
/db_xref="GI:2909576"
/db_xref="SPTREMBL:O53633"
/translation="MPARSVPRRPRAVRVAVRGLVAVMPRRKSGIPALDGLALAV
ALVASHGIGPMGGGFGVDATFVLSGFLTSLIDELGTRGLIDGFWIKRRL
LPALVLMVLTSAARALEPDDALTLRSDALIAELMTRMRFRVANTDYEFOGAPSP
LOHTSLGVEQYVWVPLLLIGATLLAARRRRCRATYGVYRFAAFILASIGTMA
SATAAVAFSAATRDRIYFGDITRAOALLISAAALLVRWPSLNRCWCLIRRMWR
RIARLFPVGLAGLAVTAVAGVEFRHGLIVAGAAVTVASVAMBERGAVARI
LAMRPLVMIGTISGVYLMHMPILFALNGRFGMSGPALEFACATVAVASWML
EOPRIRWAVRPLPLAATVAVASAAVATMLVPGVAGPGLREISLTPGVSAVAVSP
SPEASOPAPGRPDNRPTVSVFSGDSTICMTLMHLPDTPGREFDITVLCSTLVGT
PYRYIGOTLEQREDCDGPANSAOVNRODPVALLIYGRRETVDRVNEGHTHGD
TFDAYLNAELORALSTVSTGVAVMTVTVPSRGEKPDGRLTPEDQERKNKRAML
HNLSQHSNWGMIDLNKKLCPDGVYTAQVKGIRKRSDEVHLTQEGVWMLFMLEDSV
V"
5626..5629
/feature="Rv0112"
5626..5629
/feature="Rv0112", (MTV031.06), len: 318. gca, possible
dehydrogenase similar to many GDP-MANNOSE 4,6-DEHYDROGENASE
genes e.g. gplU18320(PAU18320.1 Pseudomonas aeruginosa
GDP-D-mann (523 aa) opt: 415 z-score: 513.2 E(1): 4.4e-21;
27.0% identity in 318 aa overlap. Contains Short-chain
dehydrogenases/reductases family signature PS00061."
/codon_start=1
/transl_table=11
/product="gca"
/protein_id="CAAI7306.1"
/db_xref="GI:2909577"
/db_xref="SPTREMBL:O53634"
/translation="MKVWITGAGGMSKSHLAEMILAGHDVATYTCRPTDPSDLOFN
GAEVDTDMCVSDSIATFRPDVAFHLAGSIPAVSWARVETLTNNMGCTATVFPAL
RRVRHAKIYVAGSSAEYGFVDPSEVFINERRELRPLHPYGVSKAADMLATYQHSY
GMHIVARIKNGTGRKRYGADLSDVFRRCWLEHHPQSAIRVGNLTKTKPTVDVDEL
NRALMLMKDEAGADYVNGSIAVENGDVYKQVIAACKRDIDVPEVDPALNRPDCK
IIVGCSKLAATGMOEDICLTOTIADMPDWRKSSSALMV"
6021..6107
/feature="gca"

```

```

misc_feature
/feature="PS00061 Short-chain dehydrogenases/reductases
family signature"
6024..6029
/feature="gca"
/feature="ASNI site: ATTAAT linking fragments N/V"
6667..7257
/feature="gmba"
6667..7257
/feature="gmba"
/feature="gmba"
/feature="gmba", (MTV031.07), len: 196. gmba, probable
phosphoheptose isomerase similar to many e.g. H. pylori
AE005. FASTA scores: gplAE005(HPAE00596.11 Helicobacter
pylori from bases (192 aa) opt: 451 z-score: 573.4 E(1):
1.9e-24. 45.1% identity in162 aa overlap"
/codon_start=1
/transl_table=11
/product="gmba"
/protein_id="CAAI7307.1"
/db_xref="GI:2909578"
/db_xref="SWISS-PROT:O53635"
/translation="MCTAPAEELFVETIAVKTRILINDRVLLAARATGRLIAGYRA
CARVPCNGCSAADQHPAELTGHIFRPLGAEALHANSHTAVANDYDVTY
FARLDEGSARPGDTLPATISTGMSVLRRAKARELGYTVAMTGESGQLAFERADF
LINVPSRDTGRIDQSHVTFTHALSEVHALPAPRO"
7289..7861
/feature="Rv0114"
7289..7861
Query Match 64.0% Score 16: DB 1: Length 13350;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CDS
gene
/feature="Rv0114"
/feature="Rv0114", (MTV031.08), len: 572. Possible
acetyltransferase similar to bacillus subtilis
BSAB0014_194 (634 aa). FASTA scores:
gpl299117BSUB0014_194 (634 aa) opt: 520 z-score: 531.7
E(1): 4e-22; 36.4% identity in 382 aa overlap. Tblast score
is 0.937"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0114"
/protein_id="CAAI7308.1"
/db_xref="GI:2909579"
/db_xref="SPTREMBL:O53635"
/translation="MPARSVPRRPRAVRVAVRGLVAVMPRRKSGIPALDGLALAV
ALVASHGIGPMGGGFGVDATFVLSGFLTSLIDELGTRGLIDGFWIKRRL
LPALVLMVLTSAARALEPDDALTLRSDALIAELMTRMRFRVANTDYEFOGAPSP
LOHTSLGVEQYVWVPLLLIGATLLAARRRRCRATYGVYRFAAFILASIGTMA
SATAAVAFSAATRDRIYFGDITRAOALLISAAALLVRWPSLNRCWCLIRRMWR
RIARLFPVGLAGLAVTAVAGVEFRHGLIVAGAAVTVASVAMBERGAVARI
LAMRPLVMIGTISGVYLMHMPILFALNGRFGMSGPALEFACATVAVASWML
EOPRIRWAVRPLPLAATVAVASAAVATMLVPGVAGPGLREISLTPGVSAVAVSP
SPEASOPAPGRPDNRPTVSVFSGDSTICMTLMHLPDTPGREFDITVLCSTLVGT
PYRYIGOTLEQREDCDGPANSAOVNRODPVALLIYGRRETVDRVNEGHTHGD
TFDAYLNAELORALSTVSTGVAVMTVTVPSRGEKPDGRLTPEDQERKNKRAML
HNLSQHSNWGMIDLNKKLCPDGVYTAQVKGIRKRSDEVHLTQEGVWMLFMLEDSV
V"
5626..5629
/feature="Rv0112"
5626..5629
/feature="Rv0112", (MTV031.06), len: 318. gca, possible
dehydrogenase similar to many GDP-MANNOSE 4,6-DEHYDROGENASE
genes e.g. gplU18320(PAU18320.1 Pseudomonas aeruginosa
GDP-D-mann (523 aa) opt: 415 z-score: 513.2 E(1): 4.4e-21;
27.0% identity in 318 aa overlap. Contains Short-chain
dehydrogenases/reductases family signature PS00061."
/codon_start=1
/transl_table=11
/product="gca"
/protein_id="CAAI7306.1"
/db_xref="GI:2909577"
/db_xref="SPTREMBL:O53634"
/translation="MKVWITGAGGMSKSHLAEMILAGHDVATYTCRPTDPSDLOFN
GAEVDTDMCVSDSIATFRPDVAFHLAGSIPAVSWARVETLTNNMGCTATVFPAL
RRVRHAKIYVAGSSAEYGFVDPSEVFINERRELRPLHPYGVSKAADMLATYQHSY
GMHIVARIKNGTGRKRYGADLSDVFRRCWLEHHPQSAIRVGNLTKTKPTVDVDEL
NRALMLMKDEAGADYVNGSIAVENGDVYKQVIAACKRDIDVPEVDPALNRPDCK
IIVGCSKLAATGMOEDICLTOTIADMPDWRKSSSALMV"
6021..6107
/feature="gca"

```

```

gene
/Note="Clinical strain"
Complement(139. .2397)
/gene="MT0112"
CDS
Complement(139. .2397)
/Note="similar to SP:P46840 PID:559912; identified by
sequence similarity: putative"
/codon_start=1
/transl_table=11
/product="cation-transporting ATPase, E1-E2 family"
/protein_id="AAK4334.1"
/db_xref="GI:13879590"
/transl_table=11
/translation="MAAPVVGADADLSVRRIRIDVSGSCAACASVETKLIKIGVK
ASVNEATRVATIDAVGMADELGVGEGKAGYHAAPITETTVDRKIDKDGHNARLL
RRLVAALFVPLADLSTLPAIVPSARVWGMTLITLALAPVVAAMPFHSVALRNA
RHRTSMETLISGIVAAATAMSLSSVFQDQPPREGSIRALINDSDIYLEVAAGVTV
GYLAGRFEERAKSKAGSALRALAEAGANVAVLPDGAELVIPASELKRRPFTTRP
GETIADGVVDDSSAIDMSAMTGEAKPVPAVPAASVGVGVMDGRVLYEATAGAD
TOPAMVRLVEAOQOTOKARAOGLADHAGVPAVAVVIGLAGAANLVSAGADAPAS
VTLGVLVACPCALGLATPTAMVAVSGRGAOLGIFKGYRALETTISIDVYPTKGTI
LVGOLAVSTVMAAGSTSERDEEVLGLAAVEASAPHMAAAIIVAASPDGPVNGF
VAAGCGVSGEVEGHEVQKPSMTIRTPPCDALVSRIDGESRGELVFEVSDGV
VRAALITADTKDSAAVAALRSRLITLITGDRRAADVAAGV IDSAVDMLEP
EGKVDIORLREEGHTVAAVGSDINDGPALVGADGLAIGRGTVALGAADIILVRD
LNTVPALDLARATMTITRMNIMAFVVAIP1AAAGLNP1IAGAAAFSSFFV
NSLRLENGAO"
2541. .4055
/gene="MT0113"
2541. .4055
/gene="MT0113"
/Note="similar to GB:247047 SP:P07278 PID:171159
PID:172689 PID:4303; identified by sequence similarity:
putative"
/codon_start=1
/transl_table=11
/product="cyclic nucleotide-binding protein"
/protein_id="AAK4335.1"
/db_xref="GI:13879591"
/transl_table=11
/translation="MTPVTPPLVVALIAGDRNIDGVIILAOHLLOTTHAMRLSLF
RYGIDPRVAVGKCYSTHPGVDVAMRDGTYVDCSDAVAPHESFPIQYTRHVERF
AESMRRLTGRARVALLDDGSDLAVAGMIDASADVIGIEOTSAGYATVGCALGF
PIINARSSAKLLESPILARVITQTAEPFAGIDSSAAILITAGATIGVALDIYRP
LHDRVVDYTRSGCMTPIDLPNAIGGVYVIGATGATSVAPSMHELLIPVILLMSMS
SDREFDAVALRRRTTPNDCHADRLVADGSDATLLNSGFPVNFDDSGPMDGSMALT
MALLAAVLVAVSAVADEMSSDHPHCLIDGDIYASPLINDVPLQASLPLSLDSD
YRRLVRSRGITLFRGGRADHFEVTESELEBALDGVILITLIGADHFGACLLGMR
RLATYRACEPVLNLDKARGDALHGDAMRLAYGVARTRLMHAGASESLAV"
Complement(4204. .4488)
/gene="MT0114"
Complement(4204. .4488)
/gene="MT0114"
/Note="similar to PID:1651940; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein L28"
/protein_id="AAK4336.1"
/db_xref="GI:13879592"
/transl_table=11
/translation="MSARCOITGRVYGFCAVSHSHRTRRRPNTIQKAVYLPSED
RRIKIVYSAGIIVIDRDSHGRRAARGASAPAHFARAGSSSLRTAIL"
4598. .5794
/gene="MT0115"
4598. .5794
/gene="MT0115"
/Note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK4337.1"
/db_xref="GI:13879593"
/transl_table=11
/translation="MRTPVILVAGODHDEVGALLRRGTVVVHHRPDRGVVRRMA
TISRGEITTEDEALFAGCVSCTIRDLALLRLRLHRDNGRIVLILAWLEPPI
CMAIDHVRVCGHGFDPGPAALDVRVAAVVTCVDCVRLPQSLGEDLDPGRVAGVT

```

```

gene
Complement(5889. .10766)
/gene="MT0116"
CDS
Complement(5889. .10766)
/gene="MT0116"
/Note="identified by match to protein family HNM"
/codon_start=1
/transl_table=11
/product="cation-transporting ATPase, E1-E2 family"
/protein_id="AAK4338.1"
/db_xref="GI:13879594"
/transl_table=11
/translation="MKIPVATVVGVTNGVAVOTVAGARLPGSAAVOTLASVLE
LTPGVVAVQTTGRALIVRGSHNSPGMTPRVWRSGRVRHDLPLPFRWHR
AAVEPVRPIGVAAHVEGSLGLVLEPDDADSDIAVEVDVVSAYAAOIFLG
SVSSPASFADPGRNELVLP1TAAMDIVAMGATVGVARLPAPROTTLALIL
NHQPRVSLMESRLGRVGDIALATTAANGTOSIGTFLDLVQSLQISEAAHR
IOTAKSRAGPVEQVNVNANGSLIAASALVAGGEDAAGATILAGVPRAHNGRDA
FAVUGRLANTGQLVDPGARLRIDRVVVVIGAAIRGDNRAVLHAQDDEGDD
RYTEADALLHEQAPEDPDELPTATGARVVLRAVAGTEDELSASVSTHPPGLKIVE
GSDVGEVDPYAIRLQTAHRTGARVVLRAVAGTEDELSASVSTHPPGLKIVE
LADRGVLLITVAHRDFASTDTLAALAIDVGAALDPRGATPMTADLITGDIANA
VRLSALPARASESAVHLAOGGTTLAGLLVTEGDDKTPNPSFRRLNPNVAAA
TALVSGMWSAAKVLRRPPTPOTLTAHMLADLPELVYRSLAGSSRPILAVEGITAMRT
LDLISYEPVMAPLRGPARTLAOLAVATRHIELADLPTLIVAGAAASAIVGSDILALY
AGVMVNAITGVQVORLREAAAELEFODLVRRVVVPAVATRLRELAARHARTTA
TVSASLKVGVYDIDAEVVPADARLIVAEVDVDSFTLGSLLPVQKVDVAVND
PKRMSLFEGSTVAGHARAIVAVGTGVAHNRASVAVDERTAGVQARLRELTGV
LPTLAGAAVATALLERRASIROAVDVAIDVAVAAPGELPVATLSOLAARLTA
KCALVRSPRTIEALGRVDTICFDKGTGLTERLRLVVCALPSSTAARDDLPOTDPS
AEVLAARASTOPNCEGNAHATDEALIAAASALASISQDSSEVYLAIEGLKVL
RGYAAATIGRVGTDGIPMLMKGAPETLLPCCRADPCVDYDEHAEVVRHIAEGLKVL
VAOKRTNGTTHDETDIDADAVDVAHDLEIGVGLADYARSSSRPLLEALIDEARN
KVOTIADHPITARAIARQLGIPADARVVAETAEALVADIEERHAAVLADMGVFAVSEP
KVOIVAAALDVEGSMAGVADVANTILVGNNGEVLFTVGTAFGARAPGROLIVNLT
LTFDEPALAVATISOPAEPPDAEPTDDAEARORHRAVRLGPTPSIDALPQIV
GSACVILQIOTPVTSHPFCTPLGPAVMTGVFSATAGATAVSALAPKWLASTGVVO
PDEPDDAEOSDAG"
10806. .10931
/gene="MT0116.1"
10806. .10931
/gene="MT0116.1"
/Note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK4339.1"
/db_xref="GI:13879595"
/transl_table=11
/translation="MAYORHVVYRLAHTSPGEVAGSGSPDPFGLRAMNDAVSDLA"
Complement(11120. .11326)
/gene="MT0117"
Complement(11120. .11326)
/gene="MT0117"
/Note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK4340.1"
/db_xref="GI:13879596"
/transl_table=11
/translation="MPEVTIHSGBPITTVNGCGGORYIVLESKTVGDSCVLLESRVN
HQLQVLEKSPFAGYHVGRAHNRIL"
11497. .13098
/gene="MT0118"
11497. .13098
/gene="MT0118"
/Note="similar to GB:AL123456; identified by sequence
similarity: putative"

```

```

/codon_start=1  
/transl_table=11  
/product-"PEP_RCRS family protein"  
/protein_id="AAK4341.1"  
/db_xref="GI:13879597"  
/translation="MRVGGPNPRNRNVRPKPHYIWLATROLARKORSIMSLITTS  
PATVAATAATHLAGIGSTALSTANAAAAAAPTTLASVAGADESVLLTALEFAYDTOLY  
SAQALAHPHOVOVALMGAVCYAAAEETNAFTPLQLTVVDNVLTAVNPAPLALLGP  
TITGNGANGLPNTGCCDGPGCGILLFGNGMNGSGGVPOAGANGANAIGLFPSGTGGMGCV  
GINGSAGGAAGAGAGGLIFCNGRPGGAGCTTGTDGDPGAGANGNLGRPMNGVGACAG  
GYCASGGGNGMNGNAGNGSTAGLIFHGCAGCAAGICGSADGLGGGGNGRMNGVGACAG  
GYSASGGGNGMNGCGVLGVGFDDGGAAGTCGLSDIVWGLAGTIGGNAGPFYNGCACAGAG  
LGSAVSASACGMGNCGLVFNGNGDGLGPTTSAGNCGMGNAVGLFPGGACAGACAG  
SGRGAGIPGRCGRGDGSGCLIDCGCTGGCAGAADAASAAGCNAGNARLIGNGGGGGPG-  
MFGRPGAGAGSSGTTIFCPAGTPGPS"  
13141..13995  
/gene="MT0119"  
13141..13995  
/gene="MT9013"  
/note="Identified by match to PFM protein family HMM
```

```

Query Match      64.0%; Score 16; DB 1; Length 1415;  
Best local Similarity 100.0%; Pied. No. 1.4e+02;  
Matches   16; Conservative    0; Mismatches     0; Indels       0; Gaps        0;
```

```
Oy          4 caccagatgcacgcc 19  
Db 13524 CACGAGGTCATCCG 13509  
|||||||  
|||||||
```

```
RESULT 27  
AE007184 LOCUS  
DEFINITION Mycobacterium tuberculosis CDC1551, section 270 of 280 of the complete genome.  
ACCESSION AE007184 AE000516  
VERSION AE007184.1 GI:13883785  
KEYWORDS  
SOURCE ORGANISM  
  
. Mycobacterium tuberculosis CDC1551.  
Mycobacterium tuberculosis CDC1551  
Bacteria: Firmicutes; Actinobacteria: Actinobacteridae:  
Actinomycetales: Corynebacterineae: Mycobacteriaceae:  
Mycobacterium; Mycobacterium tuberculosis complex.  
1 (bases 1 to 14634)  
REFERENCE  
AUTHORS Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O., Peterson,J.F., Nelson,W.C., Dodson,R., Gwinn,M., Holt,D., Hickey,E., Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M., Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H., Gill,J., Mikula,A. and Bishai,W.  
TITLE Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 14634)  
AUTHORS Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O., Peterson,J.F., Deboy,W.C., Dodson,R., Gwinn,M., Holt,D., Hickey,E., Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M., Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H., Gill,J., Mikula,A. and Bishai,W.
```

```
TITLE Direct Submission  
JOURNAL Submitted (22-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
```

```
FEATURES  
source location/Oualifiers  
1..14634  
/organism="Mycobacterium tuberculosis CDC1551"  
/strain="CDC1551"  
/db_xref="taxon:83331"  
/note="Clinical strain"  
120..1136  
/gene="MT3903"  
120..1136  
/gene="MT9303"  
/note="Identified by Glimmer2; putative"  
/codon_start=1
```

gene
CDS
1216. 2997
/gene="MT3904"
/note="similar to SP:P33224 GB:L20915 PID:457172
PID:457174 PID:537028; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="acyl-CoA dehydrogenase, putative"
/protein_id="AAK48270.1"
/db_xref="GI:13883787"
/translation="MEPYLEAVNDKLPSTPEKAQRYOTENYRCAGLNWYLDTPLIQ
FMAVYIRPDELAFAAPH/IRIGELTGCVTMAEETRNPLERYDROHDISRYV
LPSEFIOKRAVIEARQAVRDOARAGVPSIALPAADVLNQAIGACALAGCNL
VRSILVAPADVPREFVLAKPNSGDEGAAOLTERGGSDGLALETTRSGDVL
LNGKWPANSGAGEAFVNLAKPBGAPDSTRGVATELVRTPDGSRNCVRIIRLDKL
GRAVESGELFEVDAAEFLFSGPSADGSPDGGLTRMELTNRLRGCTAFNLGN
RRAVEISGICACGRRAFGGALIDKPLMRKRLAEWVDAVALAMVDFEGCANHROP
CLPRIVAPRTIKTCTCLGITTVASDAIEIHGNGYLETVPVRLIDKQVNTMGKD
NIIICIDYRCIETRAHETLRLRDAVSDDDPYRLVSRIIDLDAALTYMRLPD
ROLAEALPLFAQCMGQVNGALLTEQAMERATRGTDKRALVARLYARVYLDQGL
RGIDADDELQRFDELVAGATFAEQT"
3134. 4468
/gene="MT3905"
/note="similar to SP:O10621 PID:1322437 GB:AL123456
GB:AL123456; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="TS1557, transposase"
/protein_id="AAK48271.1"
/db_xref="GI:13883788"
/translation="MRNRLFRALIGYDKRTVIEDIEEEDDAGDGAARVYARVPSA
VYRRCGRGGRASWYDCRGLRQMSLDMGTEVEFLAEARVNCPTPHGPTVAAPWA
RIRHAGHYRQAVDVAWLVACSTAVCELMIAIRTVGALVARWMADETKRIDRFNL
RRTIGDIEISYKRHNRYLTVVYHDHSGRLVMAAPKDHGATGLFEDALGARRAQITHV
SADADWADADVTERCERPAIDGCAQDPPIVMAAPALDPPRRRANODAPLAEPIWG
RORCPKNAAPRGRERARRKLGARTALMKNPEDITEOSAKTKRIRVADIALDHGILSL
KSLSTENKRIETLRIGAFGRSPQVLAALMLATLAGHRTPLGRHNHPOLSQ"
complement(4521. 6089)
/gene="MT3906"
complement(4521. 6089)
/gene="MT3906"
/note="similar to SP:P53003; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="propiionyl-CoA carboxylase, beta subunit"
/protein_id="AAK48272.1"
/db_xref="GI:13883789"
/translation="MTPEPVYHTTAKEKLAEERLELAKEPGEGKAARXDKQIPSP
AARIRYELVDPGSMETIGALCRTPDDNALYCDGVYGHGILNGPVCFSHDQTVGEG
GTGVEGGRVAVRILMEWCAMVCGPTVGINDSGAILDQAVYISLAMYATLGRAPYTSNI
LVPOALISLCKCGAGVAVSPIQTDLVAAKRDGTFMVFVGPVTDVYDEPDSIDELGGS
ADHOASGNIHQVVESEAAAYOYVRDFLSPNCFKRPVAVNPOLREPTIGHDELD
SYDPSDNMAKYDEVLRLIEDGDFLDVAADQAGAILTGAARVDGRIVGVANOPHIN
MSCADINESAADKAARFIRPSDADFDILVEVVDTPGFLPGVEDEKNGIIRKGRFLYAA
VEADVVPVTTIKRSYGGAAYVMGSKOLADNFMPTARIIVADGDAADOLMKRPF
DPNAPEQADIRKSFVENYINMAIPIWIAERGRIDAVIDPHETRIILLKSHMLLRDKC

```

gene
complement(6086. .11287)
/gene="MT3907"
complement(6086. .11287)
/gene="MT3907"
/note="similar to GB:000023 pid:467207: identified by
sequence similarity: putative"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAK48273.1"
/db_xref="GI:13883790"
/transl_table="MADVAESONAPAEARELTPPEMROWLKNWYKAVGAPSIDC
SVPMEVIGLSSRDVAMADIEDLTGTVSTVAFAHPHTEISLARTIEGEPEDLAG
DDEADSRGCPARVDIALVGLSTFRPGPMNPECTONALLEGRTDLDGKRESE
LEPRLAARVAGARTGKYLKDKGDSSEFFVAKTEADNDPOQMALLETWALHE
ARIPASSIRGAGVYIGSSSTNDYSFLASDPTVAHPAATGSSIIANVSVFYDF
HGPSTITDINACSSSLVAIHQVQALKNRGADVVAGVNAITTPWTLGPEIGAVLA
PDRIRKSFADAGTIRSEGGMLVKRVDAIRGDAILAVIAGSNHDSRNGLI
APQDQADAVLRVAKDAGIDPRVDYIEAHGTGTLIDPTIEALGVRVGRPADIR
PALLGAVKTNVGHLESAGASMAKVVIALOHDKLPSTINAGSPYIDPMAMKMI
TPTDMPRYGVALAGVSSEFCGANAHVVRVLPVYKEPPEPPEPPEPPEPPEPPE
APLAGALRPDEPGNIIIDSAVAEPEPELPGYTEALRLKEALELAQEVETAL
VPLASAPLIRKKAAMAEIADMQSPGQASSLESIGRSLSRHHGRRAVLAHDH
DEKIKGLRAVAGKQAPNTFSYDGPVTTGPVAVLAGFGAHRKMGKSLYLENEFAW
IERVDALVODELIGISVLELIDDAODYGIEITTOVTIPIALIGELLRHGAKPAVT
GSLGSAASAVFAGI.SLRDARAIKCSRSHLMGBEALFGEYIRMALEYSADEIR
EYFSDPDLVCVAAPOQVIGGPEQVDAILARAEECKFAKPAKTKASHSQMD
PLIGELTAELQIGKPTSPGIGFVHEGRIYKGGPIIHVYKWKGLRHSYFTHG
IRNAVDSGHTTFLLELAMPVALMOVALTTAGLHDQALITTLARKODEVSSMTVA
OLYVYGHDIITLFSRASGPDYANITPTPEKKEKHEMLPAHSGDSITYMGTIVHAL
PGRHVMEYAPDGNVDLAALVRAAAHVLPDAOLIAEQAAGVGGDSTYMPGTIVHAL
GSAVOVARIDESTLYVDALVRSAGESEVLAIRGLVGSAMGEPDLPEPELLEL
ADAPLTSOLITRMPSGMTWSPDSCETIAERGLVGSAMGEPDLPEPELLEL
GLIDSLMAVIRKNEYEDLPILQTLAVRDANILNVEKLEVEVHEHDEEDLHEOK
TUTAEIARAQELHKGKTEPEDESEAGVALPSPNGGPPPTGALNVDPPEDA
AERTVFAWALVITKSPGIFNELPRIDDEAKALPISRSERAGPITLADVLTSNI
BALADRVYTLAAGIDGPFVTLRAREDAQKVEYFTHGSGSTVAYEELDLRIPAD
VREGLIDAVAGEITQIKELIRKMDRYAAAEKFTNTIPIAIFYOLEELDDGGO
VRPVLDAVSSGVOIPAGIIEHRTSYLDNRALDIAOIOYDGHVTLVADRYHDAAI
MEPRYAVRQPDGGMGEVSDLEVPVIGEHQIAIDPEITAKVGEHMSRALQTEADR
TSEVRGQ"
complement(11294. .13183)
/gene="MT3908"
complement(11294. .13183)
/gene="MT3908"
/note="similar to PID:1658531: identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="acyl-CoA synthase, putative"
/protein_id="AAK48274.1"
/db_xref="GI:13883791"
/transl_table="MAYHNPFTVNGKIRFPANTNIVRHEKNAKVGKLAIFYLDFS
TERDQVARDILMSDFSRNNAVGARLQOVTOPDRVAILCPONLIDYISFALYSGR
IAYLFDPAEPGVHGRHAVALDDCAPSTLITTDSEAGRKIRARSKKEPRRYAVD
ADPEVAATMOQPEANEPAVLAQYTSSTRIJDSGTIHTNLPTNVVOVINALGGE
GDGVSMLPFFHDMGLITVLASVLSGHSFTMPPAFAVRKRELRELKAFYGGGT
FSAAPNFAEEHAALVGVPRDDEPDLIDSVKQILNGSEPSASRKFEFAFAGYK
OTAVKSYGLAEATLFTVSTTPMDEVPTVIHVDRELNMGREVEAADA.PNVAASVS
KVGSRMAYIVADTRASELPDQGTGELWLNGLITGVWKEESAQFKNILKSRTS
ESRARECAPDALMVRTGDGYFKHLYIAGRIKDLVILIDRGNHPODELETAQESRK
ALRGTAAAFSVANOLDPQVYDDSHAGLKPEDEPTSEOLVIGERRAGCHIKIDHPTI
VVDIRAIIVGVHGVTVRDVLLVASTIPTRTSGKIGRACCAAVLIDGSRGVSPTV
FATISD"
complement(13496. .14506)
/gene="MT3909"
complement(13496. .14506)
/gene="MT3909"
/note="identified by Glimmer2: putative"
/codon_start=1
/transl_table=11

```

```

/product="hypothetical protein"
/protein_id="AAK48275.1"
/db_xref="GI:13883792"
/transl_table="MAKNSRRKRRHRLIAMIAGAMAVVAIVVYIMLGASPPS
AVPPGVLPPEPTPAHPKPRPAEODASCPYOMITVPCWSSROONPLINTVOPKAL
LNKVTGPIADQFAPARVONTVAYTAOFNPINLTITTONKSTYNDREAGTRAAVAMTD
MNRCPILSVYLLICFSOGAVIADVNSDIGNGDPVDEDLVLGVTLIADRGQGVNGO
VPSPPREGAEITIEHREVPVLSGLVLMGPRGSGALDGRNETICAGDILCAAPAO
AESFANLPTTLNLTAGAGAPVHAWAYATPEFWSNDSGEPTMTLMAHOLIENAPHRK
HR"
BASE COUNT      2365 a 4931 c 4768 g 2568 t 2 others
ORIGIN
Query Match      64.0%: Score 16; DB 1; Length 14634;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 cgcagcaggtgcatcg 17
|||||
Db 4576 cgcagcaggtgcatcg 4591
|||||
RESULT 28
LOCUS MT3906 23740 bp DNA BCT 03-AUG-2001
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 157/162.
ACCESSION AL022076 AL123456
VERSION AL022076.1 GI:3256026
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis H37Rv.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 23740)
AUTHORS
Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Tekle, F., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,
Connor, R., Davies, R., Devlin, K., Feldwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squires, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and
Barrell, B.G
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature. 393 (6685), 537-544 (1998)
89295987
2 (bases 1 to 23740)
REFERENCE
Parkhill, J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 26, 1998 this sequence version replaced gi:2961395.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a hidden Markov Model of 78 genes
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gty, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream

```

FEATURES
 source
 misc_feature
 RBS
 gene
 CDS
 repeat_unit
 gene
 CDS
 gene
 CDS
 gene
 CDS

initiation codon.
 Location/Qualifiers
 1..23740
 /organism="Mycobacterium tuberculosis H37Rv"
 /strain="H37Rv"
 /db_xref="taxon:83332"
 <1..23740
 /note="fragment designated v026. Does not represent a physical clone"
 60..64
 /note="possible RBS for RV3797"
 72..1853
 /gene="fad35"
 72..1853
 /gene="fad35"
 /note="RV3797, (MTV026.02), len: 593, fad35, probable acyl-CoA dehydrogenase similar to E. coli AIdb protein, gp1209151ECOAIDB_3 E. coli (aldB) gene, comp (541 aa): FASTA scores: opt: 455 z-score: 561.5 E(): 9.2e-24, 31.1% identity in 514 aa overlap. Also similar to MTC1376.2 M. tuberculosis cosmid (542 aa), AIdb, 31.9% identity in 458 aa overlap"
 /codon_start=1
 /transl_table=11
 /product="fad35"
 /protein_id="CA17861.1"
 /db_xref="GI:2950416"
 /db_xref="SPTREMBL:O53577"
 /translation="MPEYDLEAVDKLPSTPEKQRYOTENYRGAMGLNMYLDPPTIQ
 FIMAYLRDELAEAPHLTRIGELTGDPVTRMAEDTRNPLERDRKDHQHSRY
 LPESIOSKRAVTEARQAVRDARAGVAPSLAEAYLLNODIGACALATGGM
 VSIATYAVPDPVREVLKLSNGEMDEAQLITERAGSGDLALETTRASDVM
 LNCFWANSCAGAEFVVLKPEGAPDSTRGVAPLTVTRDGRNCRVTRIKDLK
 GTRSVASCEIFEVDAFAFLISGEPADGSPGKLTIRMELTNRLRGTSFALGNA
 RRAVLSTCYACORAFRFGALITDKLMRKLAENVDEALAMVDFEGANRQPR
 CLPQRIAVETPKTKTCRGLTYAASDAIEHSGNGYETIMVPARLIDQAVTMEGP
 NLCIDVRGIEQTRAHETLRLARLAAVSVSDDDPTTLVSRIEDLAATLMTKID
 RQLAEARLPELAQFMGDVYVAGALLTEQAAWERAATGTDRKALVALYARRLADGGPL
 RGTADCCDEALQFDELVAGAFABDT"
 1980..3311
 /gene="RV3798"
 /note="151557-3rd copy"
 1980..3314
 /gene="RV3798"
 /gene="RV3798"
 /note="RV3798, (MTV026.03), len: 444, Putative transposase identical to 151557 MTCY373_32, sp10106211Y02X.MYCU HYPOHERICAL 50.2 KD PROTEIN CY37 (444 aa), 99.8% identity in 444aa overlap and similar to many, e.g. gp10576491DBU57649_1 Dibenzoofuran-degrading bacterium (165 aa), FASTA scores: opt: 767 z-score: 926.8 E(): 0, 67.3% identity in 168 aa overlap, or to Rp151001.1 B. paraperlensis insertion sequence (406 aa), 24.9% identity in 402 aa overlap."
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein RV3798"
 /protein_id="CA17862.1"
 /db_xref="GI:2950417"
 /db_xref="SMISS-PROT:Q10621"
 /translation="MRNRLFLRALILGDKRTVIEDIEEEDDAGCGARVYARPRPSA
 VLRGRCGRASWYDRCAGLRBMSLDMGVEVFLAEARVNCPPHGPVAVPWA
 RHNGCHYAFDDTVAMILVAVSKTAVCELRIMARVGAIAIARWADTEGRI19FAL
 RRIAGDLSYKRNHRYLTVVVDHDSGRVMAAPGDKATGLCPDADGASRAQI17HV
 SADADWADAVTEKCPRALIOCAPRPHKVNKPEDLERRRANNDARAIARPEKGS
 RQPGKNAAPRGRARRRLKGAIRALKNPEDLERRRANNDARAIARPEKGS
 KSLRHSVSKCEEGQALDRWISMAQCR1PVEYELARIKRHRVAILDALDHGLSO
 GLIESTNTK1RLTRIATFGFRSPQALLALMLTLAHRP1LGRHNP10SO"
 complement(3382..4935)
 /gene="accD4"
 /gene="accD4"
 complement(<3382..4935)
 /gene="accD4"

/note="RV3799C, (MTV026.04c), accD4, len: 517. Probable propionyl-CoA carboxylase. Beta chain, similar to many, e.g. sp1530031PCCB_SACER PROPIONYL-CoA CARBOXYLASE BETA CH (546aa), FASTA scores: opt: 1549 z-score: 1841.5 E(): 0: 50.7% identity in 519 aa overlap. Also similar to U00012.17 M. lepreae cosmid (549 aa), 48.8% identity in 514 aa overlap and to MTCY71.20 M. tuberculosis cosmid (548 aa), 48.9% identity in 515 aa overlap"
 /codon_start=1
 /transl_table=11
 /product="accD4"
 /protein_id="CA17863.1"
 /db_xref="GI:2950418"
 /db_xref="SPTREMBL:O53578"
 /translation="MTVEPVLTHTAEKLAELRERLEAKPEGGEKAARKDKGIPS
 ARARIELVDGSPFMEIGALCRIPGDNALIKGGVYTGCLINGRVGVSHDQVFG
 LGVGEEMGRKVALRMENCAVGCIVGINDSGARIQDVAITSLAWAEIGRHHLLSG
 LVPQISIIICKAGCAGVYSPIQDLVVAVRDGVMFVTGVDYTKDVTGSDYSLDEGC
 ADHQAQYNIHQVVESEAAVYQYRDLSPNCSKDPVYVNPGLPEITGHLELD
 STVPSDMMAYDMHEVLLRIEDDGLDVAQAQGAIIINGYARVDGRTGVVANDPMM
 MSGAIDNASKARAFTRFSPADPIPLVEYVDGPGFLGYEQEKNGIIRGGRFLVAV
 VEADVPKVTITTRKSYGATVAVMGSKQTLTDMLNPMTARIAVIGADGAQLMKRFP
 DPNAPDAQAIKRSFEVENTLMAIPLWIAERGFIDAVIDHERILLRSMHLRDKO
 LMRVGRKHG"
 complement(4932..10133)
 /gene="pks13"
 complement(4932..10133)
 /gene="pks13"
 /note="RV3800C, (MTV026.05c), len: 1733. Probable polyketide synthase, similar to many, e.g. gp1636761SERERYAA_1 S. erythraea first ORF of eryA gene, involved in complex polyketide formation in erythromycin biosynthesis. FASTA scores: opt: 1329 z-score: 1411.0 E(): 0: 33.6% identity in 1232 aa overlap or to Af040570.9 Amycolatopsis mediterranea rifemycin biosynthetic gene cluster (3413 aa), 34.3% identity in 1162 aa overlap. Also highly similar to mycobacterial putative polyketide synthase: MTCY338.19, U00023.3 M. lepreae, MTCY22G10.2, MTCY19H9.2, U00023.6 M. lepreae, p500606 Beta-ketoacyl synthases active site"
 /codon_start=1
 /transl_table=11
 /product="pks13"
 /protein_id="CA17864.1"
 /db_xref="GI:2950419"
 /db_xref="SPTREMBL:O53579"
 /translation="MADVAESDENAPAEAEITVPEBMQMLPNMYKAVGAPDSIDE
 SVPEVELGSSRDVAMADIEDLTGVTLSVVAERHNPITELARITIEGSPNDIAG
 DADMDARTGPARRVDIAIVGLSTRPGEMNPEQWQNLLEGGITDLPGRNSES
 LEPRILAAVAGARTRGYLDIKGDFSEFFVAKTEADNIDPQRMALLETWELH
 ARJASSTRGQAGVYIGSTNDYSFLASDPTVAHPVYITGSSITIANRYSYDF
 HGKSVTIDTGCSSLYAIIHQVOALNGEADVAVAGVALLTPMVTTLGFDIGAVLA
 PNDGAKISFSADADIDYRSEGGMLYIKRVDARDDDALIYAGSAVHIDGRSGLI
 APNODQADVLRRAVKGADIDPRTVIDTEAKGTLPSTNFGASPYIDFDMRLKMI
 PALLGAAVTUNGLHLSAAGAASAKVLLAIDHKLPSTNFGASPYIDFDMRLKMI
 TTPIDMPRYGVALAGVYSGFCGAGNANHVVRVLRDVRVERPEPRPKAABPAE
 APTLAGALRDEFEGNIIITDSAAVEEPEPLPGVTEBALIKFALEBLAODVYAPL
 VPLVASFLTSRKKAAAEELADMMQSEGGASLESIGRLSRNRNGRSYLRNEFVAM
 DEAIKGLRAVAGKQANVFSVQDPVYTGWVALAGAQHMRMGSYLRNEFVAM
 IERVDALVDELGYSYLELIDDAODIGRTITGVTTIAIDILAGELIRHNGKRAVI
 GOSLGEASNAVFAGLSLRDTRATRCGRSHLMGEMLEGEYIRLMALEYASADIR
 EVFSDPDELVCVYAAPQVYTGPPQDVHILARABEKKPARKATGASHTSOMD
 PLIGELAEILOGIIPKTSPTGCIESTVAHGRIKGGEPHIDVYKMKKGLPASYFPHG
 IRNAVDSHTTFLLELAPVLMQVALLTDAGLHDXQLITLARKODEYSSVSTMA
 OLYVGHLDLIRTEFSRASGPQDANIIPPRFKKEMLPAHNSGDSSTVMTTHAL
 DGRHVMYUAPRQGVNDLAALVYRAAAHVLPDADQTAABORAVGDDARLVTPMTHAL
 GCAISOYHARIDESFTLYYDALVSRAGESTLAPVAGAAIIVAGDAPVAPETPAD
 ADAETLSLITTRKRYEYDDLPDITLARRONMYNVKLIIEYVENHDEVOOLHEHOK
 GIDSLMAVRIKRYEYDDLPDITLARRONMYNVKLIIEYVENHDEVOOLHEHOK
 TQTAEBIRAAQAEILLHGVGTEPEVDSAGVALSPONGDQPNPGLVAVDPPRA
 ARVTEAFWALIVTGGKSPGICFNEPLRDEDEAAKIAQRLSRAGCPTTADVLTSSNT
 FALADKVRVYLEAGOIDFVRLRARPAGCKVYVFFHFGAGGSTVVEYELGRLPAD
 TMVYGFERYEGSIIEBBAQYVPKLIEMQGDPPVYLVMGSLGGIYALVACAIGLRLRGLD

VRFVGLIDAVRAGEEIPOTKEEIRKRMDRVAFAPAKTENVYIAPITPEQIEEDIDBQ
VRFVGLIDAVRAGEEIPOTKEEIRKRMDRVAFAPAKTENVYIAPITPEQIEEDIDBQ
MEPRYAVRQDGGGGEVSDLEVPVIGETITQIDEPITAKGHEMSRACQIEADR
TSEVCKO"

complement(4945, .4948)
/gene="pk513"

Query Match 64.0%; Score 16; DB 1; Length 23740;
Best Local Similarity 100.0%; Pred. No. 13e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
2 ccgacagagtgatcg 17
|||||
Db 3422 CCGACAGAGTGATCG 3437

RESULT 29

LOCUS
POL245436/c

DEFINITION POL245436 34902 bp DNA HCV 09-MAY-2001
Pseudomonas putida OCT plasmid alk genes cluster (alkBFGHJKL, alkN
and alkST genes) and flanking DNA, strain TPA-1L (GPO1), formerly
annotated as Pseudomonas oleovorans.

ACCESSION
VERSION AJ245436 J04618 J04619 S50571 X52935 X65936
AJ245436.1 GI:5824135
KEYWORDS alkane-1 monooxygenase; alcohol dehydrogenase; aldehyde dehydrogenase;
alk gene; alk gene; alk gene; alk gene; alk gene; alk gene;
alks gene; alk gene; hypothetical protein; methyl-accepting
chemotaxis protein; ORF; outer membrane protein; plasmid; jca
gene; reca gene product; rubredoxin 1; rubredoxin 2; rubredoxin
reductase; tnpal gene; transposase.

SOURCE
ORGANISM Pseudomonas putida.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.

REFERENCE
AUTHORS 1 (bases 1 to 34902)
Kok, M., Oldenhuis, R., van der Linden, M.P., Raaijmakers, P., Kingma, J.,
van Leliveld, P.H., and Wiltholt, B.

TITLE The Pseudomonas oleovorans alkane hydroxylase gene. Sequence and
expression

J. Biol. Chem. 264 (10), 5435-5441 (1989)
89174581
2 (bases 1 to 34902)
Kok, M., Oldenhuis, R., van der Linden, M.P., Meulenbergh, C.H.,
Kingma, J., and Wiltholt, B.

TITLE The Pseudomonas oleovorans alkBAC operon encodes two structurally
related rubredoxins and an aldehyde dehydrogenase

J. Biol. Chem. 264 (10), 5442-5451 (1989)
89174582
3 (bases 1 to 34902)
Egink, G., Engel, H., Vriend, G., Terpstra, P., and Wiltholt, B.

TITLE Rubredoxin reductase of Pseudomonas oleovorans. Structural
relationship to other flavoprotein oxidoreductases based on one NAD
and two FAD fingerprints

J. Mol. Biol. 212 (1), 135-142 (1990)
90204534
4 (bases 1 to 34902)
van Belien, J.B., Egink, G., Enequist, H., Bos, R., and Wiltholt, B.

TITLE DNA sequence determination and functional characterization of the
OCT-plasmid-encoded alkyl genes of Pseudomonas oleovorans

Mol. Microbiol. 6 (21), 3121-3136 (1992)
93086421
5 (bases 1 to 34902)
Panke, S., Meyer, A., Huber, C.M., Wiltholt, B., and Wubbols, M.G.

TITLE An alkane-responsive expression system for the production of fine
chemicals

Appl. Environ. Microbiol. 65 (9), 2324-2332 (1999)
10347009
6 (bases 1 to 34902)
van Belien, J.B., Panke, S., Lucchini, S., Franchini, A.G.,
Rothlisberger, M., and Wiltholt, B.

TITLE Analysis of Pseudomonas putida alkane-degradation gene clusters and
flanking insertion sequences: evolution and regulation of the alk

genes
Microbiology 147 (Pt 6), 1621-1630 (2001)
21286873
7 (bases 1 to 34902)
Van Belien, J.B.
Direct Submission
Submitted (21-JUL-1999) van Belien J.B., Institute of
Biotechnology, ETH, ETH-Hoenggerberg, Zuerich CH-8093, Switzerland
On May 15, 2001 this sequence version replaced gi:261146 gi:3127910
gi:49078.

Related sequences: AccNo M83932 (IS53), AccNo AF052751 (IS1384),
AccNo U84154 (IS1491).

FEATURES
source
location/Qualifiers
join(1..15660,16825..34902)
/organism="Pseudomonas putida"
/plasmid="OCT"
/strain="TPA-1L (GPO1)"
/db_xref="taxon:303"
/focus
/note="formerly annotated as Pseudomonas oleovorans TPA-1L
(GPO1)"
212..751
/note="ORF1"
/codon_start=1
/transl_table=1
/product="hypothetical protein, 20.4 kD"
/protein_id="CA54043.1"
/db_xref="GI:5824136"
/db_xref="SPTREMBL:Q9R9V6"
/translation="MSRTSKETSNAKIOEKQANTSLITKILIRKALIEDRQRLPSL
DLKSKRGIPIFRSALMSDADLVISCSYNSRPNTEYSQDLAALFTYNNLPAT
OTLPQAKRTTORSQOEISTLKNQVDVLTNTLGEVRYAVQGLVARDEHRODIRYQO
VKSHTALDRHLTLVKP"
748..1671
/note="ORF2"
/codon_start=1
/transl_table=1
/product="hypothetical protein, 35.2 kD"
/protein_id="CA54044.1"
/db_xref="GI:5824137"
/db_xref="SPTREMBL:Q9R9V5"
/translation="MIHVAKAPRLDFEFGHTTPSLPMLPDSKNGECWEINSLTSS
IGGANSYGVADPAIKTFESHADPLNPOKIEENKLTLDVDTEDLYDEVKYKNSRT
LHSTTKIKIVRIALLIEHOKCKKQNLITDPRANQINATEDPFLRNRPRLR
LTHCDELIDPTPKPTSPTRNNELAMHEALHDYNSQYIIDRWYIATLLEHGSR
IEEENINPASTINATKNGNPVNIPLVKGKYGCLQVLIIPTELQSTNSFYSLP
ALITASPTGYSLAKILIDILMRAPLPIPTT"
2396..2755
/note="ORF3"
/codon_start=1
/transl_table=1
/product="hypothetical protein, 15.3 kD"
/protein_id="CA54045.1"
/db_xref="GI:5824138"
/db_xref="SPTREMBL:Q9R9V4"
/translation="MMRPDAKQEVYLVKPYDFKSLDGLAALVELDIKAVPPVL
EVLNLSNRKVLILYERNGFCMLMKRLESEFSPPTPEVALIYVQELNMLDGF
DLMRNPHOVLPRVVA"
2440..4351
/note="IS66 related-like sequence"
2774..4306
/note="ORF4"
/codon_start=1
/transl_table=1
/product="hypothetical protein, 57.8 kD"
/protein_id="CA54046.1"
/db_xref="GI:5824139"
/db_xref="SPTREMBL:Q9R9V3"
/translation="WISMPENPDDPVLLKOLEOMTHERASDKGKIVHLEEVALLR
OILFRKTEQIGDADTQPLPDEASLAELDADEEVAPLTKRGRKPKPLPALP
RLEVNHLEPEHLELCACGRNHAIGEVSDLEVPVIGETITQIDEPITAKGHEMSRACQIEADR
VTAADPQAIMIEKMSAPSVLAMLITTYVNGIPLHREKVLGRHGIDIPROTILARVI
OCSEHFOPLMLMRESLNSRIHICDETROYVLKEGREDPSOSMMWVOTGPPDRPV

CDS

CDS

CDS

misc_feature


```

gene
<3627..>144
/gene="T21P5.2"
CDS
join(3627..3661,3762..3843,3918..3995,4081..4140,
4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,
5361..5494,5576..5682,5785..5827,5903..6144)
/gene="T21P5.2"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAFO1580.1"
/db_xref="GI:6017097"
/translation="MDDNRSSSEIKRHETEKDTIARKEPTNTKLIQDPEMALIYK
VRSOEIHSILOERIAACIKMOLLNREYGLERKADLRVAITDRONSYSALNEL
ARRKGDLEMLKLAHDKVDEDERYFPMSTSLGLAEIGVWRVANNAVLSIGLHL
DOLPKAKACNDRIRELSTIVNOPPTDTSKONHDPNRKSTQASGSTRONDYOTN
EOLLPRENTNRDYNINODTSLKFNQIGSGSQGLPPQKRENFYPLSSVAGKE
MIOERPEKANSMPDANGNEFASVHVEBPQIDGQIIGDAIPGKRVGCGCPVR
CTTLCMORVRLHEDGTROYIRGATHPEYIVTADVDKLIAVECIPMDQGRQYKTR
FSGIYSFNSVSKDVLIMQELVRLFANDNKIRCGVFNGT"
misc_feature
6225..6289
/note="exon predicted by xgrail, quality marginal"
6384..6503
/note="exon predicted by xgrail, quality excellent"
6794..6839
/note="exon predicted by xgrail, quality excellent"
6928..7033
/note="exon predicted by xgrail, quality excellent"
7293..7411
/note="exon predicted by xgrail, quality marginal"
repeat_region
/rpt_family="(TAAA)n"
/note="exon predicted by xgrail, quality marginal"
complement(7488..7547)
/note="exon predicted by xgrail, quality excellent"
repeat_region
complement(7654..7784)
/rpt_family="(TAA)n"
complement(7819..7864)
/note="exon predicted by xgrail, quality good,shadowexon"
8054..8113
/note="exon predicted by xgrail, quality good,shadowexon"
8263..8380
/rpt_family="(TAAA)n"
complement(8592..8627)
/note="exon predicted by xgrail, quality marginal"
misc_feature
marginal_shadowexon"
<8652..>9527
/gene="T21P5.3"
<8652..>9527
/gene="T21P5.3"
/note="contains zinc finger motif, C3HC4 type (RING
finger).9527
/gene="T21P5.3"
/note="unknown protein"
/codon_start=1
/protein_id="AAFO1602.1"
/db_xref="GI:6017119"
/translation="MGSTGNPNPWGTTVDYSYDCSQGCVSYCPQWCVIFPPPEPY
LDDEDSSSDSPDLALIGILASAFILFVYLLISKYCHRRHNSSTSAALNRI
VCSLYWGTGNNNNNGATNPNOTIGGGGDDPSLITVYVYTRKDDGVESDIDS
QIATFENNSLSTGSDVVYVNDLNSRSRNETVNGSTPIGXGDFRRRRRAELGVYK
LRRRCIDRRYTKGD"
misc_feature
complement(10838..10896)
/note="exon predicted by xgrail, quality"
marginal_shadowexon"
11174..11254
/note="exon predicted by xgrail, quality good"
11690..11700
/note="exon predicted by xgrail, quality marginal"
join(<11814..12256,12381..12578,12676..13085,13162..13332,
13418..>13761)
/gene="T21P5.4"
<11814..>13761
/gene="T21P5.4"
gene

```

```

CDS
/note="similar to unknown protein GB:ANC32238 [Arabidopsis
thaliana]"
join(11814..12256,12381..12578,12676..13085,13162..13332,
13418..13761)
/gene="T21P5.4"
/note="unknown protein"
/codon_start=1
/protein_id="AAFO1581.1"
/db_xref="GI:6017098"
/translation="MAETKSGSESPYKTIYVVOENKSPFDITGLWKEINLEIDGYM
KSPQKNPSPSSNLSHNYPPDOSQYVDPNGHSIRIYIOVGKPMDSQHPDNG
PAMSGFPAQNAERKKGMSAVMNGFKPDALPVYKELVQVAGICDRMFASVPGATON
RIETHATSHGTTNNRKLILGFPQKTIIESLDEAGFTFGIYQCFPTTIFYNLK
KLTILFHHYGFQFKKDCESGLMPYVVEEDRWPTLDLNPANDHPSHDSECKLYK
EYVEALRSSPQNNELFIITYDHEGFPVPLDGVNPDGILGPPVNFERNLIG
VAPPEFISPMWIEPTGLHSGNPGPYLMSQYHSSIPATVKKIFRLKPLTRDSWAGT
FESVITRNSPDDCETLSNPYKMGVAKRNAELSDQBELVTAAGLGDVKNEL
LYKCKRTCSVASRYVYKADPKFVEESKAREKRGDENDIVFCVDDDDHNVKPP
SOSPSHATPMNS"
misc_feature
complement(14488..14563)
/note="exon predicted by xgrail, quality
excellent,shadowexon"
14893..14935
/note="exon predicted by xgrail, quality marginal"
join(<15141..15586,15675..15872,16258..16667,16780..16950,
17036..>17427)
Query Match 64.0%; Score 16; DB 8; Length 82289;
Best Local Similarity 100.0%; Pred. No. 1,le+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 aggtgcatgcacga 23
Db 30302 AGGTGCATGCCACCA 30317
RESULT 33
AC011732
LOCUS
DEFINITION Homo sapiens clone RP11-17011, WORKING DRAFT SEQUENCE, 14 unordered
pieces.
AC011732
AC011732.4 GI:7107960
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 115955)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-17011
Unpublished
2 (bases 1 to 115955)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Bana,N., Beckerly,R., Boguski,K., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collins,A.,
Cooke,P., DeBellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Galagan,J., Gargyua,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., McKernan,K., McDonald,P., Marquis,N.,
McEwan,P., McGuff,A., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirelli,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W., Zimmer,A. and Zody,M.
Direct Submission
Submitted (13-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 28, 2000 this sequence version replaced gi:524206.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

```

```

http://ftp.genome.washington.edu/RW/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3676
Center clone name: 17.0.11
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 91165 bases at least Q40
Consensus quality: 102606 bases at least Q30
Consensus quality: 109983 bases at least Q20
Insert size: 87000; agarose-1p
Insert size: 114655; sum-of-contrigs
Quality coverage: 6.0 in Q20 bases; sum-of-contrigs
Quality coverage: 4.6 in Q20 bases; sum-of-contrigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1      1835: contrig of 1835 bp in length
*
1836 1935: gap of 100 bp
*
1936 3722: contrig of 1787 bp in length
*
3723 3822: gap of 100 bp
*
3823 6826: contrig of 3004 bp in length
*
6827 6926: gap of 100 bp
*
6927 12743: contrig of 5817 bp in length
*
12744 12843: gap of 100 bp
*
12844 19565: contrig of 6722 bp in length
*
19566 19665: gap of 100 bp
*
19666 25342: contrig of 5677 bp in length
*
25343 25442: gap of 100 bp
*
25443 35236: contrig of 9794 bp in length
*
35237 35336: gap of 100 bp
*
35337 45543: contrig of 10207 bp in length
*
45544 45643: gap of 100 bp
*
45644 57815: contrig of 12172 bp in length
*
57816 57915: gap of 100 bp
*
57916 73012: contrig of 15097 bp in length
*
73013 73112: gap of 100 bp
*
73113 87110: contrig of 13998 bp in length
*
87111 87210: gap of 100 bp
*
87211 95720: contrig of 8510 bp in length
*
95721 95820: gap of 100 bp
*
95821 102140: contrig of 6320 bp in length
*
102141 102240: gap of 100 bp
*
102241 115955: contrig of 13715 bp in length.
Location/Qualifiers
1. 115955
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-17011"
/clone_lib="RPC1-11 Human Male BAC"
1. 1835
/note="assembly_fragment"
1936 3722
/note="assembly_fragment"
1936 3722
/note="assembly_fragment"
clone_end:17
vector_side:left"
misc_feature
3823 6826
/note="assembly_fragment"
6927 12743
/note="assembly_fragment"
12844 19565
misc_feature

```

```

misc_feature /note="assembly_fragment"
19666 25342
/note="assembly_fragment"
25443 35236
/note="assembly_fragment"
35337 45543
/note="assembly_fragment"
45644 57815
/note="assembly_fragment"
57916 73012
/note="assembly_fragment"
73113 87110
/note="assembly_fragment"
clone_end:sp6
vector_side:left"
misc_feature 87211 95720
/note="assembly_fragment"
95821 102140
/note="assembly_fragment"
102241 115955
/note="assembly_fragment"
misc_feature 102241 115955
/note="assembly_fragment"
BASE COUNT 31603 a 25627 c 26285 g 31102 t 1338 others
ORIGIN

```

```

Query Match 64.0%; Score 16; DB 2; Length 115955;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 gcagcaggtcgcacgc 18
|||||
Db 106529 GCAGCAGTGCATCGC 106544

```

```

RESULT 34
AC002457/c 177380 bp DNA PRI 21-DEC-1999
LOCUS Homo sapiens BAC clone CTB-60P12 from 7q21, complete sequence.
DEFINITION AC002457
ACCESSION AC002457
VERSION AC002457.1 GI:3947433
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE Smith,A., Beck,C. and Gibson,A.
1 (bases 1 to 177380)
TITLE The sequence of Homo sapiens BAC clone CTB-60P12
JOURNAL Unpublished
2 (bases 1 to 177380)
AUTHORS Waterston,R.
REFERENCE Direct Submission
Submitted (20-AUG-1997) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
3 (bases 1 to 177380)
AUTHORS Waterston,R.
REFERENCE Direct Submission
Submitted (02-DEC-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 177380)
AUTHORS Waterston,R.
REFERENCE Direct Submission
Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Dec 2, 1998 this sequence version replaced gi:2337872.
COMMENT
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_RC060P12

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7/>, send <mailto:regreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTB-60P12 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pBelorAC11

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is CTB-60N22, 200 bp overlap; the clone sequenced to the right is CTB-137N13, 200 bp overlap. Actual start of this clone is at base position 1 of CTB-60P12; actual end is at 177380 of CTB-60P12.

FEATURES

source

Location/Qualifiers

1..177380

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7q21"

/clone="CTB-60P12"

/clone_lib="CTB-978SK-B"

/complement(<131..112890)

/gene="WUGSC:H_RG060P12.2"

/complement(join(<131..215,1834..1923,31059..31144,41160..41293,90777..90892,112769..112890))

/gene="WUGSC:H_RG060P12.2"

/note="Rap2 interacting protein; similar to U73941 (P1D:G1916018); H_RG060P12.2"

/codon_start=1

/evidence=not_experimental

/protein_id="AAC82532.1"

/db_xref="GI:3947434"

/translation="MASRSUGSGCIRGGGGGKKSISANNAVERNNLTTCRFVKTLLDRSGFERIDSSPEFNNFAILLDQLSHRLKGVWYGVSPSPFDYIVACRKYVSONCISINEMVSSRAKRAWTRVALMERKHLSEYISTALRDKTKTRFEEDGAIVLGEANMLAGMLGNATIDFPCLKGGHGSFPAVIDTPLYKIQY"

480..791

/rpt_family="Alu"

repeat_region

4518..4824

/rpt_family="Alu"

repeat_region

4825..4937

/rpt_family="(TA)n"

repeat_region

5732..6004

/rpt_family="L2"

repeat_region

6012..6047

/rpt_family="L2"

6143..6231

/rpt_family="MIR"

repeat_region

30382..30669

repeat_region 6232..7254
 repeat_region /rpt_family="L2"
 7272..7412
 repeat_region /rpt_family="L1"
 7420..7813
 repeat_region /rpt_family="L1"
 7814..8150
 repeat_region /rpt_family="MERL-type"
 8151..8265
 repeat_region /rpt_family="L1"
 8284..9183
 repeat_region /rpt_family="Retroviral"
 9314..10441
 repeat_region /rpt_family="L1"
 10437..12545
 repeat_region /rpt_family="L1"
 12928..13352
 repeat_region /rpt_family="MaLR"
 14327..14443
 repeat_region /rpt_family="L1"
 14443..16612
 repeat_region /rpt_family="L1"
 16719..16804
 repeat_region /rpt_family="AT-rich"
 17093..17116
 repeat_region /rpt_family="AT-rich"
 17492..17774
 repeat_region /rpt_family="Alu"
 18097..18135
 repeat_region /rpt_family="AT-rich"
 19255..19594
 repeat_region /rpt_family="L1"
 19900..19946
 repeat_region /rpt_family="(GA)n"
 19966..20245
 repeat_region /rpt_family="L1"
 20545..20614
 repeat_region /rpt_family="AT-rich"
 20684..20816
 repeat_region /rpt_family="AT-rich"
 20817..21123
 repeat_region /rpt_family="Alu"
 21124..21134
 repeat_region /rpt_family="AT-rich"
 21136..21171
 repeat_region /rpt_family="(CA)n"
 21478..21940
 repeat_region /rpt_family="L1"
 22310..23003
 repeat_region /rpt_family="L1"
 23135..23653
 repeat_region /rpt_family="MER2_g"
 23658..24140
 repeat_region /rpt_family="L1"
 24263..24304
 repeat_region /rpt_family="AT-rich"
 24419..24447
 repeat_region /rpt_family="AT-rich"
 24618..24865
 repeat_region /rpt_family="Alu"
 25627..25650
 repeat_region /rpt_family="(TTAA)n"
 25906..26152
 repeat_region /rpt_family="L2"
 26153..26536
 repeat_region /rpt_family="AT-rich"
 26554..27556
 repeat_region /rpt_family="Retroviral"
 28842..28903
 repeat_region /rpt_family="(CAT)n"
 29888..30080
 repeat_region /rpt_family="MERL-type"
 30382..30669

```

repeat_region      /rpt_family="Alu"
                    32790..32960
repeat_region      /rpt_family="MIR"
                    33062..33192
repeat_region      /rpt_family="L1"
                    33199..33998
repeat_region      /rpt_family="L1"
                    34000..34042
repeat_region      /rpt_family="(TAAAA)n"
                    34043..34316
repeat_region      /rpt_family="Alu"
                    34414..34602
repeat_region      /rpt_family="L1"
                    35064..35263
repeat_region      /rpt_family="L1"
                    36255..36571
repeat_region      /rpt_family="Alu"

```

```

Query Match      64.0%; Score 16; DB 9; Length 177380;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      8 aggtgcatccagca 23
        1111111111111111
Db      48179 AGTGCATCCAGCA 48164

```

```

RESULT 35
AC084136/C      180700 bp      DNA      HTG      14-FEB-2001
LOCUS      Homo sapiens clone RP13-506N4, WORKING DRAFT SEQUENCE, 32 unordered
DEFINITION      pieces.
ACCESSION      AC084136
VERSION      AC084136.2 GI:12830241
KEYWORDS      HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens

```

```

REFERENCE      Mammalia: Eutheria; Primates: Catarrhini; Hominoidea: Homo.
AUTHORS      Birren, B., Linton, L., Nusbaum, C. and Lander, E.
JOURNAL      Unpublished

```

```

REFERENCE      2 (bases 1 to 180700)
AUTHORS      Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
JOURNAL      Anderson, S., Barina, N., Bastien, V., Beda, F., Boguslavsky, L.,
AUTHORS      Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
JOURNAL      Choquet, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P.,
AUTHORS      Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P.,
JOURNAL      Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M.,
AUTHORS      Graham, L., Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L.,
JOURNAL      Iliev, I., Johnson, R., Jones, C., Kann, L., Levine, R., Liu, C., Liu, G.,
AUTHORS      Lamas, R., Landers, T., Lehotzky, J., Levine, R., McEwan, P., McKenna, K.,
JOURNAL      Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McKenna, K.,
AUTHORS      McPheters, R., Meldrum, J., Menus, L., Mihova, T., Miengo, V.,
JOURNAL      Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,
AUTHORS      O'Donnell, P., O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K.,
JOURNAL      Pierre, N., Pisan, C., Roy, A., Santos, R., Schauer, A., Severy, P.,
AUTHORS      Rogov, P., Roulman, D., Roy, A., Santos, R., Schauer, A., Severy, P.,
JOURNAL      Sougniez, C., Spencer, B., Sclange-Thomann, N., Stojanovic, N.,
AUTHORS      Strauss, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J.,
JOURNAL      Tiselli, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, K., Vo, A.,
AUTHORS      Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
JOURNAL      Zimmer, A. and Zody, M.

```

```

TITLE      Direct Submission
JOURNAL      Submitted (13-OCT-2000) Whitehead Institute/MIT Center for Genome
COMMENT      Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Feb 14, 2001 this sequence version replaced g1:10800269.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

Center: Whitehead Institute/ MIT Center for Genome Research

```

Center code: W1BR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L11406
Center clone name: 506_N4
----- Summary Statistics -----
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.96071
Consensus quality: 167122 bases at least Q40
Consensus quality: 174060 bases at least Q30
Consensus quality: 176531 bases at least Q20
Insert size: 183000; agarose-fp
Insert size: 177600; sum-of-contigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 3.6 in Q20 bases; sum-of-contigs

```

* NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 951: contig of 951 bp in length
1 952 1051: gap of 100 bp
1 1052 1051: contig of 1013 bp in length
1 2065 2164: gap of 100 bp
1 2165 3416: contig of 1252 bp in length
1 3417 3516: gap of 100 bp
1 3517 4816: contig of 1300 bp in length
1 4817 4916: gap of 100 bp
1 4917 6297: contig of 1381 bp in length
1 6298 6397: gap of 100 bp
1 6398 9144: contig of 2747 bp in length
1 9145 9244: gap of 100 bp
1 9245 11578: contig of 2334 bp in length
1 11579 11678: gap of 100 bp
1 11679 30905: contig of 19227 bp in length
1 30906 31005: gap of 100 bp
1 31006 32606: contig of 1601 bp in length
1 32607 32706: gap of 100 bp
1 32707 34256: contig of 1550 bp in length
1 34257 34356: gap of 100 bp
1 34357 37618: contig of 3262 bp in length
1 37619 37718: gap of 100 bp
1 37719 39131: contig of 1413 bp in length
1 39132 39231: gap of 100 bp
1 39232 41291: contig of 2060 bp in length
1 41292 41391: gap of 100 bp
1 41392 43172: contig of 1781 bp in length
1 43173 43272: gap of 100 bp
1 43273 45886: contig of 2614 bp in length
1 45887 45986: gap of 100 bp
1 45987 49186: contig of 3200 bp in length
1 49187 49286: gap of 100 bp
1 49287 52781: contig of 3495 bp in length
1 52782 52881: gap of 100 bp
1 52882 57288: contig of 4407 bp in length
1 57289 57388: gap of 100 bp
1 57389 62317: contig of 4929 bp in length
1 62318 62417: gap of 100 bp
1 62418 65388: contig of 2971 bp in length
1 65389 65488: gap of 100 bp
1 65489 70061: contig of 4573 bp in length
1 70062 70161: gap of 100 bp
1 70162 75771: contig of 5610 bp in length
1 75772 75871: gap of 100 bp
1 75872 83423: contig of 7552 bp in length
1 83424 83523: gap of 100 bp
1 83524 89716: contig of 6193 bp in length

```

```

* 89717 89816: gap of 100 bp
* 89817 97026: contig of 7210 bp in length
* 97027 97126: gap of 100 bp
* 97127 104654: contig of 7528 bp in length
* 104655 104754: gap of 100 bp
* 104755 114340: contig of 9586 bp in length
* 114341 114440: gap of 100 bp
* 114441 126202: contig of 11762 bp in length
* 126203 126302: gap of 100 bp
* 126303 142208: contig of 15906 bp in length
* 142209 142308: gap of 100 bp
* 142309 159777: contig of 17469 bp in length
* 159778 159877: gap of 100 bp
* 159878 178832: contig of 18955 bp in length
* 178833 178932: gap of 100 bp
* 178933 180700: contig of 1768 bp in length.

```

FEATURES

```

source
  1..180700
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="RP13-506N4"
    /clone_lib="RPC1-13 Human Female BAC 481-768"
  1..951
    /note="assembly-fragment"
    clone_end:SP6
    vector_side:left
  1052..2064
    /note="assembly-fragment"
  2165..3416
    /note="assembly-fragment"
  3517..4816
    /note="assembly-fragment"
  4917..6297
    /note="assembly-fragment"
  6398..9144
    /note="assembly-fragment"
  9245..11578
    /note="assembly-fragment"
  11679..30905
    /note="assembly-fragment"
  31006..32606
    /note="assembly-fragment"
  32707..34256
    /note="assembly-fragment"
  34357..37618
    /note="assembly-fragment"
  37719..39131
    /note="assembly-fragment"
  39232..41291
    /note="assembly-fragment"
  41392..43172
    /note="assembly-fragment"
  43273..45886
    /note="assembly-fragment"
  45987..49186
    /note="assembly-fragment"
  49287..52781
    /note="assembly-fragment"
  52882..57288
    /note="assembly-fragment"
  57389..62317
    /note="assembly-fragment"
  62418..65388
    /note="assembly-fragment"
  65489..70061
    /note="assembly-fragment"
  70162..75771
    /note="assembly-fragment"
  75872..83423
    /note="assembly-fragment"
  83524..89716
    /note="assembly-fragment"
  89817..97026

```

```

Query Match      64.0%; Score 16; DB 2; Length 180700;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 8 agtgcgtccgcagca 23
|||||
Db 25268 AGGTGCATCCGACGA 25253

```

```

RESULT 36
AE003424
LOCUS
DEFINITION
  AE003424 289090 bp DNA
ACCESSION
  AE003424 AE002566
VERSION
  AE003424.1 GI:7290316
KEYWORDS
  HTG.
SOURCE
  fruit fly.
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
  Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 289090)
  Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
  Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
  George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
  Sutton,G.G., Wortman,D.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
  Brandon,R.C., Rogers,Y.H., Blazer,R.G., Chame,M., Pfeiffer,B.D.,
  Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gaber
  Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
  Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
  Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
  Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
  Botchan,M.R., Bouck,J., Brokstein,P., Brotlier,P., Burris,K.C.,
  Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
  Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
  Pablos,B., Delcher,A., Deng,Z., Drenth,J., Drenth,S.M.,
  Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
  Fleischmann,W., Foster,C., Gabriellian,A.E., Gay,N.S.,
  Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
  Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
  Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
  Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
  Kienison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
  Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
  Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Matzel,B., McIntosh,T.C.,
  McLeod,M.P., Mohrson,D., Morkulov,G., Milshina,N.V., Mobarry,C.,
  Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
  Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
  Nusskern,D.R., Paclet,J.M., Palazolo,M., Pittman,G.S., Pan,S.,
  Pollard,J., Puri,V., Reese,M.G., Reiter,K., Remington,K.,
  Saunders,R.D., Scheeler,F., Shen,H., Shie,B.C., Siden-Kiamos,I.,
  Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
  Stapleton,M., Strong,R., Sun,E., Svrtk,R., Tector,C., Turner,R.,
  Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
  Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
  Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
  Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
  Zhong,R.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
  Gibbs,R.A., Myers,E.W., Rubin,G.M., and Venter,J.C.
  The genome sequence of Drosophila melanogaster

```

TITLE

JOURNAL Science 287 (5461), 2185-2195 (2000)
 MEDLINE 20196006
 REFERENCE 2 (bases 1 to 289090)
 AUTHORS Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
 TITLE Direct Submission
 JOURNAL Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 FEATURES
 source Location/Qualifiers
 1. 289090
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="X"
 complement(join(<9373..9459,9519..9720,9784..10373,
 10559..11005,11151..11813,12630..12711,13406..13521,
 13582..14921,15552..16167,16573..>16587))
 /gene="EG:BACH48C10.4"
 /product="CT33608"
 /db_xref="FLYBASE:FBan0014047"
 /db_xref="FLYBASE:FBgn0040390"
 complement(<9373..>16587)
 /gene="EG:BACH48C10.4"
 /note="CG14047"
 /map="2F5-2F5"
 /db_xref="FLYBASE:FBan0014047"
 /db_xref="FLYBASE:FBgn0040390"
 complement(join(9373..9459,9519..9720,9784..10373,
 10559..11005,11151..11813,12630..12711,13406..13521,
 13582..14921,15552..16167,16573..16587))
 /gene="EG:BACH48C10.4 gene product"
 /note="start-1"
 /db_xref="FLYBASE:FBan0014047"
 /db_xref="FLYBASE:FBgn0040390"
 /protein_id="AAF45777.1"
 /db_xref="GI:7290317"
 /translation="MDSILAVGAGGVVAAQGHPRVQVYTFYTLTELFQQRNNS
 /HGVATGATVDRDTRDSHGHNGHGHGHEQJLELLIDCKRIKNTKPKSSALIL
 AWMKQGLDKOQOQARLAATARSQENYSAEDELIPNSDSEDERITVYRQOLEKT
 KRLNLAKENGGLSARPNFGKRLSVEELMSATSDIYSTSEAGVTSVIRKATDSVG
 AMPNRSRSGECNPVITGSLRLHLDPPVSVSLNCHGCSAAYCGSGPKPLPPRTYV
 VAQMCQEPREPANQOAKAIVITETARKMRSSPKSGSGSATGSGVSGSPKRYA
 CHRCQPEPETHAOLSPDKMEHOTCLIESPKQTTTLEOKQEDQLSLMLIGLAQLA
 PAATLCQGERIKENSTPTIAVVPPTDSVLTSTTHWDSGSSNGGTTSTSTA
 TTTTKQPRALITENPEDSCDESPIDEPPYRPMSSALRFGTMSLEKIPSDRME
 ADELDDDDLSKFNCHDGNSPNEEDDDADSEPALYONDLGASSGCVINQGLAS
 GANTNAGAYVSDKMSFEESKAFIDKILGRNAGDAQOQHOQCTASPTDOMECTS
 GATSGEEVWGTPSGDNDQDLQINSEHNSPTKSTSLNDDDTDELMMDLILMA
 PPTASTIRGLPRRLLEPFEETSEDEKTOQSDDIKGEATTNGHYLEDSAGSS
 SDSEVPAPPLVPRPVREDDEPSAEQOENLAMEDQEDLHSSVERLLPNTATSAV
 GRRLLTTLRLPTTAIDPPPAVLVASECVLLDORPSSVGRDLIAGSICITMTTS
 RPTPLSPAKAAATPTGMATMTTNTTTTMTTPSPRAPSTTSTISSADITSSGGRAS
 RPPRVVPPPPPRRLLLTQTLIAGQAKTWLLGGCRRRORRPPNRRVADNAINTE
 YQOLVSGSPKTTVINDNKRDLPAGGQALSTFINSLKNGKADPPSTHAPVAEG
 SELPGCSASTIDGCARCTPAPNPMRPRPKPAGNAGQOPHSGRGLDLIRKMSR
 NDYSGMDPTVNRPRPLSAGALGSGTSSNNPNSRSHNSHNSLGGGLPAGYEV
 GQOOLVVOAGGSGAGRVGSGSAGSAGTSSNKGSLGLDLILARERIKYSQS
 RSCGRVRETTSTTTLMAMATEMRSGMDSEKRDSSWPKRAAASLITRSTSTMDGAG
 AISTTTTTTTTTTTEAFLETFEVRSKRLINFLIKRNSSETTASTSTSTNFSGAIGEO
 PPSHSQSSLLLOKLPFGQSQADSEVMAQMSPKDKNPKSLNRLMKQIKRRRTNS
 VSOIVAG"
 complement(join(<18933..20031,20093..20270,20329..21208,
 21574..22759,24311..24418,25054..25330,25668..26541))
 /gene="EG:BACH7M4.1"
 /product="CT33604"
 /db_xref="FLYBASE:FBan0014045"
 /db_xref="FLYBASE:FBgn0040387"
 complement(<18933..>26541)
 /gene="EG:BACH7M4.1"
 /note="CG14045"
 /map="2F5-2F5"
 /db_xref="FLYBASE:FBan0014045"
 /db_xref="FLYBASE:FBgn0040387"
 complement(join(18945..20031,20093..20270,20329..21208,
 21574..22759,24311..24418,25054..25330,25668..26541))

21574..22759,24311..24418,25054..25330,25668..26541))
 /gene="EG:BACH7M4.1 gene product"
 /note="EG:BACH7M4.1"
 /codon_start=1
 /db_xref="FLYBASE:FBan0014045"
 /db_xref="FLYBASE:FBgn0040387"
 /protein_id="AAF45778.1"
 /db_xref="GI:7290318"
 /translation="MPPTMRMRHSSSSAVVEESGRRRGVGVGVDANKENGVHF
 /translational="MPPTMRMRHSSSSAVVEESGRRRGVGVGVDANKENGVHF
 MSSPFGNASTLALODLSNVHGSPPORRSRSGSGSPROATPOLAALRCIPPTGGAVL
 EDSQSSSRNGDTTLDRLDAITESARKVYRQTKLPAGATTTTLADNAEWSERS
 VHEMEVYRPTHLKROVVRKKNPHTTTTNOQSQAQKLEPDLVPTKRCISFSSS
 SASSDDEDEQVAKRSLSAPTPPSPCTTTSISSSSNGGADMEASQSGSD
 VSIAPAKBOULNVHITCRDLORSNGSINAYVALVSGAGPVGYSGGGS
 WOKRVASGTPRTAPKLANAKSEELGCTEPFLNLMVHMDRLNSVPQYQSE
 MOARVSGTPRTAPKLANAKSEELGCTEPFLNLMVHMDRLNSVPQYQSE
 SSGSSNSAVALNDEYTSISIDSMKEDLLPQOORNNHNNCTSGRROSTMPNNGS
 FLIEDPPADNANSTTQAOATGSSASAKANSHNNHNNCTSGRROSTMPNNGS
 SVGAVVQOQGRPTMTKRLRTEREGFSTIWHHPVREKIEAGLSADRCGLTLC
 DYVIFVDKHNVTMPADVNLIRSGSSILTEIFRSGAGATTITSDGONVHIS
 TRCAVAGLSEETLATTATTGTVMSLQRTTSTRIOPLANSRSPATACGTVTSI
 EAKRRRLHPMTPEKSEIVPVNDNRRLLOLSI SEONFTALHFGVDRFVQJGER
 KDLISPDHRTLQNDILRIAPDLLEDCSSDODORPOMFASRYLSKTAICA
 AYKRYCNGIKRADVLVAKRSQTSERFIATPPAPRRKRPDLTMFIHRLDHFEL
 KIMQILAGNCHVDTEHNKSTVINELQAYRSPVWTLFSDIYFAKVSRYLFI
 RMVPTKCPPTLAVQGRMTFGDDLSEVRSKVPWTLFSDIYFAKVSRYLFI
 TEEPTPLANVQSCFHKRKTTERFLVNDPNSGLASPGYCAPDLITPLSDIGLTA
 LILRAPSLERKAWONLQROIPLVNAALSTPLSPDLPDVLNPLVSLDGLTA
 SMGSMKLPDLSIHLKQOQOYVLFESKRLDGSVENLQRLASAVRHHCASSTSS
 ATQTOHPOPLCTTTSIPRNSPARNRQPNRQNGVSGVPGVGVGFGGCTNNS
 CLSSASLTQTOVOTLNLHSGTSHGQVSLTTEPTPTPNTSPPTSPNSDAENFV
 SQKLEFSLSSVGRSFIDESGVGLLELDQLHEVNDPESTDYDVECLFLDEDERECTE
 AISTTGKCPNIPLAFFGSDWLELDQLHEVNDPESTDYDVECLFLDEDERECTE
 GHGHDEDEDESVVDDPFGMLPTRPDPDSLTL"
 complement(<53568..>54629)
 /gene="EG:BACH7M4.4"
 /product="CT33234"
 /db_xref="FLYBASE:FBan0012496"
 /db_xref="FLYBASE:FBgn0040385"
 /evidence=not_experimental
 complement(<53568..>54629)
 /gene="EG:BACH7M4.4"
 /note="CG12496"
 /map="2B6-3A1"
 /db_xref="FLYBASE:FBan0012496"
 /db_xref="FLYBASE:FBgn0040385"
 /evidence=not_experimental
 complement(53568..54629)
 /gene="EG:BACH7M4.4 gene product"
 /note="EG:BACH7M4.4"
 /codon_start=1
 /db_xref="FLYBASE:FBan0012496"
 /db_xref="FLYBASE:FBgn0040385"
 /evidence=not_experimental
 /protein_id="AAF45779.1"
 /db_xref="GI:7290319"
 /translation="MKWTPASQRLNEIPVHDODANOPORRLQRRORRATANK
 ENPVDTVGTSVTPITRLRNSPVLAPLSIRNTPPAVYRSQAPQVQSVRATL
 PRPAEYSPNGLPSGQHLRGLAPLDPLQGPRIYVSGAGYNNLQRRDLDEPVAT
 LEPELOETAVPEPPKRVPRSTPQAPSTQGDQDLDLDAIDVPLCKANASSKKK
 PRSTRTLRRTLVQOMESNCPQEVLSVAGDPPSDLSFGVLVPLMNLTPPE
 PASPLAKIPNLIVOLATVPAPASPCDTCLETTPVVALKTRKEIVAKSPILKRYK
 VDERNVEFGALPSSTIV"
 complement(join(<70555..71849,71925..>71976))
 /gene="gt"
 /note="Nucleotide sequence of the Celera sequence differs
 from the published sequence for this transcript."
 /product="CT33972"
 /db_xref="FLYBASE:FBan0007952"
 /db_xref="FLYBASE:FBgn0001150"
 complement(<70555..>71976)
 /gene="gt"

```

/note="CG7952"
/db_xref="FLYBASE:FBan0007952"

Query Match      64.0%; Score 16; DB 3; Length 289090;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gtagcgcgcgcgcgc 25
    |||||||
Db 13743 GTGCATCGCCAGCATT 13758

RESULT 37
AF374832/C 1185 bp DNA PLN 27-MAY-2001
LOCUS Saxifraga fortunei 26S ribosomal RNA gene, partial sequence.
DEFINITION AF374832
ACCESSION AF374832.1 GI:14210120
KEYWORDS
SOURCE
ORGANISM Saxifraga fortunei.
            Saxifraga fortunei.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Saxifragales; Saxifragaceae; Saxifraga.
            1 (bases 1 to 1185)
            Soltis,D.E., Kuzoff,R.K., Mort,M.M., Zanis,M., Fishbein,M.,
            Huffor,L., Koontz,J. and Arroyo,M.
            Elucidating deep-level phylogenetic relationships in Saxifragaceae
            Ann. Mo. Bot. Gard. (2001) in press
            2 (bases 1 to 1185)
            Soltis,D.E., Kuzoff,R.K., Mort,M.M., Zanis,M., Fishbein,M.,
            Huffor,L., Koontz,J. and Arroyo,M.
            Direct Submission
            Submitted (25-APR-2001) Department of Integrative Biology,
            Washington State University, Science Hall, Pullman, WA 99164, USA
FEATURES
    source
        1..1185
            /organism="Saxifraga fortunei"
            /db_xref="taxon:159961"
            <!.>.1185
            /product="26S ribosomal RNA"
BASE COUNT 279 a 273 c 381 g 252 t
ORIGIN

Query Match      60.0%; Score 15; DB 8; Length 1185;
Best Local Similarity 100.0%; Pred. No. 6.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 aggtgcgcgcgcgc 22
    |||||||
Db 441 AGGTGCATCGCCAGC 427

RESULT 38
AF219383 3069 bp mRNA INV 05-JAN-2000
LOCUS Drosophila melanogaster triptop protein mRNA, partial cds.
DEFINITION AF219383
ACCESSION AF219383.1 GI:6671378
KEYWORDS
SOURCE
ORGANISM Drosophila melanogaster.
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
            1 (bases 1 to 3069)
            Fasano,L., Nhouyvanitsvong,M., Cault,X., Kerridge,S., Jacq,B. and
            Volta,C.
            Teashirt-related paralog
            Unpublished
TITLE
JOURNAL

```

```

REFERENCE
AUTHORS 2 (bases 1 to 3069)
          Fasano,L., Nhouyvanitsvong,M., Cault,X., Kerridge,S., Jacq,B. and
          Volta,C.
TITLE Direct Submission
JOURNAL Submitted (28-DEC-1999) LGPD, CNRS Case 907, Parc Scientifique de
          Luminy, Marseille cedex 9 13288, France
FEATURES
    source
        1..3069
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /chromosome="2L"
            /map="40B-40C"
            1..>3069
            /note="zinc finger protein: similar to teashirt; genomic
            DNA deposited under GenBank Accession Number AC006415;
            expressed in the Malpighian tubules and stomatogastric
            nervous system"
            /product="triptop protein"
            /protein_id="AAP23183.1"
            /db_xref="GI:6671379"
            /translation="MLHEAVMLEITRQALSAELTSPRCQSRDNTSAGAGAGADYR
            CPNESHCSANDRTTPAATPTLTPEATISPSVGLPLTATLPAAAVALLPPOSAM
            AAVLAQGNHLLTNPLAAASLVQHTQOAVGVEESPALPSPRRKPSHGDDO
            EEDQEDQDEQDEBDHVDQDNGPDLISYGRORSVPAPAKTIPRTSADYKP
            LPPGSMWPIINPYLAAVAAKTGLGYSKLAPSEASKALEKTEMSELETSTARRSG
            ATSSVAGVAGAGASSNSGGRHSAMQSHLNKAGDPAKDYFCWCOKQSFSTLALTLH
            MKETORGVQIPSPITPGVCGTPSPPTRLATASNSACSSSSSTSSNSKSHL
            NMLIKEMMLPRKIVRGDVGILKAGDTRILKCMWGQSFSLAMTSHMETQHY
            TNLISQELTISKSDERKRPNTGVPTSTAAPSPSCTAPSVAVLTKVQDAQG
            SIKEIETHMAQSHYKESPPAPSPAGTNPARGRONREKRLSLPVKXLEIER
            SGNSSLSDLAKPLRDFAAATKITECKGSKIEFALVEHTRKCLGESITPIPRNSA
            GVDRLPSLGLGAEKPPSVNALLEOLEKEFERTRTHTGGYSAGPILASITLK
            RLGIEDSDYTPPLMDAQAMHLRSSFSGRSCSSSSSSSTSSSTSSSTSSSTSS
            SDTPAPPPEPPPTIKAEPLAEPLVPGVDRGCSPPROIOYKKEFSMEKRESPPVS
            KSPAPOTRSPRSDNGSILALNSMDQSGVNSGNNNSGCFNNNSCSSLVSAQKPA
            HPLALQIKICTPTPPSTGLASASAGSSTASATLPSANGNDLVAFSMAQCEVLSAS
            NGSAGSDSITIKCTDTPFASKATVRHHSKVFHVDAGDESDPRLSKSPAVQSPRSM
            LASPRRSRSRPSATSOOPPSPTIPSDSPSKFLYELAKQLSKSKNA"
BASE COUNT 761 a 917 c 803 g 588 t
ORIGIN

Query Match      60.0%; Score 15; DB 3; Length 3069;
Best Local Similarity 100.0%; Pred. No. 5.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 agcaggtgcgcgc 19
    |||||||
Db 2114 AGCAGGTGCATCGCC 2110

RESULT 39
AF274670 3369 bp DNA PLN 13-AUG-2000
LOCUS Tetracentron sinense 26S ribosomal RNA gene, complete sequence.
DEFINITION AF274670
ACCESSION AF274670.1 GI:9799468
KEYWORDS
SOURCE
ORGANISM Tetracentron sinense.
            Tetracentron sinense.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; Trochodendrales;
            Trochodendraceae; Tetracentron.
            1 (bases 1 to 3369)
            Fishbein,M., Hibsch-Jetter,C., Soltis,D.E. and Hufford,L.
            Phylogeny of Saxifragales (Angiosperms, Eudicots): Analysis of a
            Rapid Ancient Radiation
            Unpublished
            2 (bases 1 to 3369)
            Fishbein,M., Hibsch-Jetter,C., Soltis,D.E. and Hufford,L.
            Direct Submission
            Submitted (30-MAY-2000) School of Biological Sciences, Washington

```

State University, Pullman, WA 99164-4236, USA

FEATURES
Location/Qualifiers
1. .3369
/organism="Tetracentron sinense"
/specimen_voucher="Qiu 94166 (IND)"
/db_xref="taxon:13715"
rRNA
1. .3369
/product="26S ribosomal RNA"
BASE COUNT 806 a 824 c 1049 g 689 t 1 others
ORIGIN

Query Match 60.0%: Score 15; DB 8; Length 3369;
Best Local Similarity 100.0%: Pred. No. 5.7e+02;
Matches 15: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 aggtgcatcgccagc 22
|||||
Db 441 AGGTGATCGCCAGC 427

RESULT 40
AF274671 3369 bp DNA PLN 13-AUG-2000
LOCUS Trochodendron aralioides 26S ribosomal RNA gene, complete sequence.
DEFINITION AF274671
ACCESSION AF274671.1 GI:9799469
VERSION
KEYWORDS
SOURCE Trochodendron aralioides.
ORGANISM Trochodendron aralioides.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Trochodendrales;
Trochodendraceae; Trochodendron.
1 (bases 1 to 3369)
Fishbein, M., Hibsch-Jetter, C., Solitis, D. E. and Hufford, L.
Phylogeny of Saxifragales (Angiosperms, Eudicots): Analysis of a
Rapid, Ancient Radiation
Unpublished
JOURNAL 2 (bases 1 to 3369)
REFERENCE Fishbein, M., Hibsch-Jetter, C., Solitis, D. E. and Hufford, L.
AUTHORS Direct Submission
TITLE Submitted (30-MAY-2000) School of Biological Sciences, Washington
JOURNAL State University, Pullman, WA 99164-4236, USA
FEATURES
source
1. .3369
/organism="Trochodendron aralioides"
/specimen_voucher="Qiu 94157 (IND)"
/db_xref="taxon:4407"
rRNA
1. .3369
/product="26S ribosomal RNA"
BASE COUNT 801 a 816 c 1060 g 690 t 2 others
ORIGIN

Query Match 60.0%: Score 15; DB 8; Length 3369;
Best Local Similarity 100.0%: Pred. No. 5.7e+02;
Matches 15: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 aggtgcatcgccagc 22
|||||
Db 441 AGGTGATCGCCAGC 427

RESULT 41
AE007551 11684 bp DNA BCT 27-JUL-2001
LOCUS Clostridium acetobutylicum ATCC824 section 39 of 356 of the
DEFINITION complete genome.
ACCESSION AE007551 AE001437
VERSION AE007551.1 GI:15023209
KEYWORDS
SOURCE Clostridium acetobutylicum.
ORGANISM Clostridium acetobutylicum

Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium.
1 (bases 1 to 11684)
Nolling, J., Breton, G., Omelchenko, M. V., Markarova, K. S., Zeng, Q.,
Gibson, R., Lee, H. M., Dubois, J., Qiu, D., Hill, J., Wolfe, P.,
Tatusov, R. L., Sabathe, F., Doucette-Stamm, L., Soucaille, P.,
Daly, M. J., Bennett, G. N., Koonin, E. V. and Smith, D. R.
Genome Sequence and Comparative Analysis of the Solvent-Producing
Bacterium Clostridium acetobutylicum
J. Bacteriol. 183 (16), 4823-4838 (2001)
JOURNAL
MEDLINE 11466286
PUBMED 21359325
REFERENCE 2 (bases 1 to 11684)
AUTHORS Childress, D., Zeng, Q. and Smith, D. R.
TITLE Direct Submission
JOURNAL Submitted (24-JUL-2001) GTC Sequencing Center Production,
Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100
Beaver Street, Waltham, MA 02453-8443, USA
location/Qualifiers
1. .11684
/organism="Clostridium acetobutylicum"
/strain="ATCC 824"
/db_xref="taxon:1488"
/db_xref="ATCC:824"
158. .565
/gene="CAC0362"
158. .565
/gene="CAC0362"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK78342.1"
/db_xref="GI:15023210"
/translation="MYIDNLRQHTIYSSIKMLGKIDKMSDEYDAEMAKTVSLA
GTLKYLQSEDKFLYPELLKNKDEATISMAKEYIDGDNIGKRFEEYKKNYKRIKVC
DNVLSLKQDTEVLNLVLRNLRNKKEDKALYPLIN"
700. .1404
/gene="CAC0363"
700. .1404
/gene="CAC0363"
/codon_start=1
/transl_table=11
/product="Deoxyinosine 3'-endonuclease"
/protein_id="AAK78343.1"
/db_xref="GI:15023211"
/translation="MKTAVHGFENASKEEFQVIOSSLVKRIKLSDFKEDIDLCAG
VDLAYWTGEGKQGVCCIIYDNTNGEIIERAYDGEIEVYMGFLAFRLPLVIT
VKTKNEPDIEMHGDNGYLYHNMGIATHASFLNKPITGVAKSYLWAGVDFPMPES
FEGAFDYINIEVYCRVLRKTDVKKPIFVSCGNIDLETCTKICSKLINDSRPLIT
VRLADLETNRKSELS"
1423. .1926
/gene="CAC0364"
1423. .1926
/gene="CAC0364"
/codon_start=1
/transl_table=11
/product="Metal-dependent hydrolases of the
metallobeta-lactamase superfamily"
/protein_id="AAK78344.1"
/db_xref="GI:15023212"
/translation="MFDPIGNGSAPFTEGRGNNSAFIKKNSMILVDCGTYPHRIKOL
NLIDQIKNLYITITTHPHDHSIGSLDLYFCYIILFKPITIIYPMKDIITDFENLIG
GDEMVELYSDNKNVLTAPFEGDINIEFIASHVGIPAYGFLGMSDNLVYVSGDGNK
LPERNK"
2267. .3214
/gene="CAC0365"
2267. .3214
/gene="CAC0365"
/codon_start=1
/transl_table=11
/product="Phosphoglycerate dehydrogenase"
/protein_id="AAK78345.1"
/db_xref="GI:15023213"

gene
CDS

```

/translation="MKTKALFTHFSEDEIKNLEGEFELLIRREKDLKYNLELEVE
EMFTYMPREHEDIAKLRKLMIOLESDGEFOIPKRYTKNNGLIVTNKNGSVSPISW
LWVSLITSTKEANIEFPOKODKRWKMLRTSLREIFETVGIIGTGHIIEATKRLQATC
VNVGLITNGKREVEFNECYAKDRVEEMIKEDGVVSLPLTIDEMHHFIRNRTILFKR
NGVAVNARSRGVVEEKLIFELKNGKIRAAALDVFEFEPIKSDSPIMONNVAITLPH
NSWISONAKRRRYKLVEENMRRYKAGBOLKNIIVLOKCY"
3311..4216
/gene="CAC0366"
3311..4216
/gene="CAC0366"
/codon_start=1
/translation="CAC0366"
/product="Predicted permease"
/protein_id="AAK78346.1"
/db_xref="GI:15023214"
/translation="MYVEFOVILFIMGYVAASKIIDDDTSHLSNEIVSITLP
LMTISFNEVERSKRTVITITNLVFSVAFTISIVGKIISFKFAMDKRIDLPMFSIF
SNCGTFPPVLKVVYGNKGVLVYSTFNLVYNVFIWIGIIVINDKREKIDYKRTLFNH
NITAVIVGVELMLSLIKIPYVMSAFNLISGMTAPLSMIVIGSLIAGVDNDIFKDM
LYVIALRLIITLPIIVFALKFOINKIVIGVILICEAMPGSLCPLLAKSCNRNKY
ASKIVLITITLSMTIPFMTLFLKI"
4240..5889
/gene="CAC0367"
/note="rocb"
4240..5889
/gene="CAC0367"
/codon_start=1
/translation="CAC0367"
/product="Arginine degradation protein (gene rocb)"
/protein_id="AAK78347.1"
/db_xref="GI:15023215"
/translation="MSRISIEIENLTVELVWPSVNGTEGEGNIGKILSYLEIDYF
KHEEYAPRIPLKNDLGRGNVPALIRGEKGSNRTIILGHMDTVAVDDYSLKDYA
EDDITLOKRLKELEIPEVEVEDLDGDMFEGASPMKSGDAVHLVLELSKRYDI
SGNITLMSNVEENENHNGIREALDLEKREYSLEITAIINDYICPMYKCDNRYL
YGTGVLKLPAPFVYKETHVQCEGEGIADVIASVINEIDNLNDLDEVNGETILP
PSYLFKRDLESYNVOTPESSAFYNYVTHSSPROASRLKAKKARFENVIRKNDI
RYKKECENTNOKYEPILNFOAKYITVOLTAKARSGDLDEIKSYEORLDVNDVJ
REISLRYVDKVNINGEKRPCTIMFAPPYCCHNLTNKNKEKEIKINKECAOOLK
EREKINEPKNVQFPRLSDSYLKIIDDSEGISLSTNFGMDRVGVVPOKIKDLN
IPAINGCTGKAKHAKWTERKYPYSFEVLPLKLMLEKFL"
5956..7242
/gene="CAC0368"
5956..7242
/gene="CAC0368"
/translation="CAC0368"
/codon_start=1
/translation="CAC0368"
/product="4 animobutyrate aminotransferase"
/protein_id="AAK78348.1"
/db_xref="GI:15023216"
/translation="MKMSILERSLKVIPVAGRATKGVYRGEGATLYTEDGKVLVD
FASGVAVCNLGHNDPAVIAKAEQMDKLIHGCHVYVYESVLAELIVELTGKRTW
YFSNAGAEANGAIAKIAKTYKROAIIISFKSGHRLATTSVKSXSKRYKNKNGLL
PYSVFAEYPCGRCPIKONKSCNMECISQFEDMFKRIIEBSVAALIMEVQEGGL
IYPPKRLKAYREICDRTIGLIPDEVQCGRGKLIIEHNEVEVEDITTCAKAALS
GPIISAYIGKELEMKWPAAGHGTFCGNPACASAIATIKELSGVLDNMMNGNTL
KEBLIKLKDYACIGDIRGIGIMGEFCENNDPDITITREAVANNLILGCG
TEHNVLRFIAPLITKEISIDMAISIVEKIVEYLK"
complement(7288..7800)
/gene="CAC0369"
complement(7288..7800)
/gene="CAC0369"
/codon_start=1
/translation="CAC0369"
/product="Predicted acetyltransferase"
/protein_id="AAK78349.1"
/db_xref="GI:15023217"
/translation="MKKIIIESRILLRPLSPDLLINIEKGTINLIKVHINONALPDE
TKAISKKTAKMEKVPKDIHEWYTWLVVDKNTKSSIGFIEGEGTDEGGEVGYSI
EFNYVLDDIH"
8033..11533
/gene="CAC0370"
8033..11533
/gene="CAC0370"

```

```

CDS
8033..11533
/gene="CAC0370"
/codon_start=1
/translation="CAC0370"
/product="Cation efflux system protein, AcrB/AcrE/AcrD
family"
/protein_id="AAK78350.1"
/db_xref="GI:15023218"
/translation="MNRITKFSKNNMVEFIIILITVGGYSTKTIMESMPNINP
VYVAVIVPGASPEQVADISRPOKALISGIDGVTDVTSNENSVYVVAQEDYSAM
DKAQKEDVAVNRKIFPEESAKKSTISRISSMGAPIMTYSINSTSIDETITVNDKTE
PKLSGISGVSVEMOGTSNDIYVVDNDKILKNGLTISDVKTALQGNISIPAGSVN
VDSSTLPIEMTKKLYTTADIEAMPYAVIPNOTKYVGEAFQVQGMPLGQAVOLGO
NNGOLGOAVQMGCGMCGMCGVAGNTOSIALNOMQAQSOILLIEQVLTNPSSAO
DKTKAORTTTAAGMLQOAOQOQDLKMLQAQOLESKALSSAGSLKTAAPSGTAKSS
APSLGSSSSSSSSNASQIKVYVILKQVAVTGSADKSYFIRSNKSNIIINITYTDG
NTVNAKDVASALEDLQKSNNGIKENISDSKRYKQSDVNGMVEGVGALFATVIA
FELKNIKATIIAIVSIPLSIALILPRFGITINISLSCMAVAVGRINDSIVIE
NHRITLKDVEKGIIOVADEVSAITSTSTVTVAVFELAMISGVKVPFPAV
TVVICILASLVAATVVPVMSRIMTINKPKPEKGEAIVGFAKVLGALNHPAAVM
GVSVLPFLSLFLIKGICITOPLPSSSVNLGTTITAPGTSAEKTEALFEKYLIN
RSDVKTVYSSIGDNSSSSSSAKMQGNSGKITVILADGVNNDKSADEIIRKATDMS
KNTRKATVVOFTSGSKDNVEIVVGNMMDITAAASKATEALGKVELTNVNTSS
KPEISVAIDSKAKANEGLSPMIAAGVQGMNNTTIOGKNVNYVILGYDNKDI
NSLKVKTIOLOGASGFNLSDIADVOADPVEELNGNOVASITADVKTDTQRA
SKNAMEKIKSFKIGIPNGVTFPONGSAKSISSPFGMGMMAVAVFWYIYVAFQSP
KAPFALIFSLPFAIGAIVALFETROPLGIDGLGMLIGITYVTAIVLIDRVQNR
KGMIVREALIEAGAIIMRPYFMATITVVALTPLAAGFSEGAIVISGLIIVIGIY
VSTIIITIVVMVSVLERD"
BASE COUNT 4344 a 1445 c 2316 g 3579 t
ORIGIN

```

```

Query Match 60.0%; Score 15; DB 1; Length 11684;
Best Local Similarity 100.0%; Pred. 4.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 7 caggtgcacgcgcg 21
Db 8190 CAGGTGCATGCCAG 8204
|||||

```

```

RESULT 42
MTCY9C4
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis H37Rv.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 15916)
Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
Harris,D., Gordon,S.V., Eigmeier,K., Gars,S., Barry III,C.E.,
Tekala,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
Connor,R., Davies,R., Devlin,K., Feltham,T., Gentles,S.,
Hamlin,N., Holtzoy,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
Squares,S., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S. and
Barrall,B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature 393 (6685), 537-544 (1998)
9825967
2 (bases 1 to 15916)
Parkhill,J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome

```

1298. .2008
/a0000="Pr3558"

```
/translation="MGIORAVLLIADIGGTYNMMHNRKHLHAQWVAQLLESV  
AKGMLAKLEGDAAFWAPGQHCPCGMRPAPADAPEVPHARADQKPSRL"
```


VLAVAVAAVATGATAGLVAALSDRATATNTSGSSSESVDITMPAAGIIVFELGI
ALFALVLTMHAGILTGCLIDADGKPVITAFEPFRNLGLVLTGLLVALTFIGGLC
VLPGLIFGVAVQFAVAVNDSTSPIDSVKASITTSNGSGSVLSWLAOTFVAVLGE
LLCFVGMGLIGIIPVAVLHVYVRKISGGVVEAVPAPVPGWPGPOLA"
/complement(4465, .4788)
/gene="MT2637.1"
complement(4465, .4788)
/gene="MT2637.1"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46950.1"
/db_xref="GI:13882381"
/translation="MAAGSVDCIDHASDOISELSNVHIGRLISPTILEVIYVIM
ESHCPRAFQAPVANSGERVTNPAGDQGPASGADYASWPGGQPTGAGRTASTWP
PDSFR"
/gene="MT2638"
5008..5646
/gene="MT2638"
5008..5646
/gene="MT2638"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46951.1"
/db_xref="GI:13882382"
/translation="MAOLLESVIDAKGKMLKLEDAFAFWAPGNTSVLCGRPPQ
MROFRTREPOIKKHPCDCCKSCORDNLSIFEVAHEQKRNVELAGVDVIL
VHRKLNEVVSLEYLFMTDVVNOCLDESVKRLATPLTHDFEGIGESPTFIDLASDM
PPAVPDHSFGLIMADKFEHMLPYLLGFKKACAGFRLSGATGEEPAEMQ"
5789..6838
/gene="MT2639"
5789..6838
/gene="MT2639"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK46952.1"
/db_xref="GI:13882383"
/translation="MLPALRDVOMRRRLVIAIVSTGLVFAMTIVLTGLVGRVFA
ERTVDSMGVAFVYKAGAGPFLGSPFAQIDLPQVAPRAGVLAAPLATAPSTIIG
TSARNVAFAGPHEGPGMPRSGRAPSTDEVAVSSTLGRNLGDDLOVCAATLRIAG
IVPSTLAKIPNIFITTEGLOOAVNCOPTISSIGIDGMPOLOPOGYVNRDAVS
DLMPLKVAADATIVAVLMTIVAAALIVGSVVISALEPRLDRFAVKAGVPRSTLA
GLAQAIVVAVLAAGVGLSLIPLAPLPMVVPVLSAFVALPAIATVIGLLASVACL
RRVVAIDPALAFGSP"
6841..7833
/gene="MT2640"
6841..7833
/gene="MT2640"
/note="similar to GB:U00096.SP:P75831.PID:1787105;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ABC transporter, ATP-binding protein"
/protein_id="AAK46953.1"
/db_xref="GI:13882384"
/translation="WGLTISDLVVERSSGAYAPRIDGLSLDVAAPGSVITLGPSC
GKTTLSCLGILTRPKSGIKFDDVDITTLGGAALAKIRKQVIVTOARNLVSLTA
LENVWPLRAAGVSAARAKRAEDLLIRVNGERMRHPPGDSGGQDRAVAAIAL
DPOILADEPTAHDFIOVEEYLRILIRSLAGDRVAVVATDSMPLADLELMAA
QVSPNQPETVYKAGVLEFEGSTMDLIIVASDEEPIVELADGCEIVAKTAPCD
YFGEIGVLEHLPRSATVARSADATVGTAAQAFRRRLGVTAVADLIEHRELASE"
8110..9861
/gene="MT2641"
8110..9861
/gene="MT2641"
/note="identified by match to PFAM protein family HM
PF01173"
/codon_start=1
/transl_table=11

Query Match 60.0%; Score 15; DB 1; Length 17988;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 6 gcaggtgcatcgcca 20
DB 768 GCAGGTGCATCGCCA 782
RESULT 44
AF065159
LOCUS
DEFINITION
AF065159 35209 bp DNA BCT 03-JAN-2000
Bradyrhizobium japonicum putative arylsulfatase (arsa), putative
soluble lytic transglycosylase precursor (slta),
dihydrodipicolinate synthase (dapA), MscL (mscL), SmpB (smpB), BcpB
(bcpB), RnpO (rnpO), RelA/SpoT homolog (relA), Pdx (pdx), and
acyl carrier protein synthase Acps (acps) genes, complete cds;
prokaryotic type I signal peptidase Stpf (stpf) gene, stpf-sps
allele, complete cds; Rnae III (rnc), GTP-binding protein Era
(era), hypothetical protein, reco-like protein, topoisomerase II
(gyra), putative ABC transporter, putative Gac-oxido-reductase,
hypothetical protein, putative cytochrome C binding protein (cytC),
putative polytopic integral membrane protein, superoxide dismutase
SodA-like protein, putative inner membrane protein, Atm-like
protein (atm), adenylate cyclases-like protein, and putative inner
membrane protein (spt) genes, complete cds; and unknown genes.
AF065159
AF065159.3 GI:6655022
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Bradyrhizobium japonicum.
Bradyrhizobium japonicum
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Bradyrhizobium group; Bradyrhizobium.
REFERENCE
AUTHORS
TITLE
1 (bases 12071 to 15486)
Baird, A. and Muller, P.
A second gene for type I signal peptidase in Bradyrhizobium
japonicum, stpf, is located near genes involved in RNA processing
and cell division
Mol. Gen. Genet. 260 (4), 346-356 (1998)
JOURNAL
MEDLINE
99086246
REFERENCE
AUTHORS
TITLE
2 (bases 1 to 12375)
Mueller, P. and Stangel, D.
Extended DNA sequencing in the upstream region of stpf in
Bradyrhizobium japonicum
Bradyrhizobium japonicum
JOURNAL
REFERENCE
AUTHORS
TITLE
3 (bases 12071 to 15486)
Baird, A. and Mueller, P.
Direct Submission
Submitted (13-MAY-1998) Biology, Phillips Universitaet Marburg,
Karl-von-Frisch-Strasse, Marburg 35032, Germany
REFERENCE
AUTHORS
TITLE
4 (bases 1 to 35209)
Mueller, P. and Stangel, D.
Direct Submission
Submitted (27-OCT-1999) Biology, Phillips Universitaet Marburg,

/product="cyclic nucleotide-binding protein"

/protein_id="AAK46954.1"

/db_xref="GI:13882385"

/translation="MTTARRRRRRRGTGDAARLARNPILADIDDEOLEPLATTVERRH

VPANOMLEHAGEPADSIYIDVSGRFVAVAEAGHFAFMASGDSITGDGVIAGAAKASG

VRALRDGVYVIAAETFTMDLEATPLLOSAMLRMAQLQSPAPATARRPVIGVS

KVLVYADGACADIMRHYVASQSDRLVVLVDQRPDAVDSLATQRPVHLITCLAEEDP

SWMDRLAEVSHHPANSDFGALARRIRGASRTIAGSDPLDDDTIPISALTREGVRD

IDRFGGTSSGATASAFALGMDAGDIAAAEFITAGSDPLDDDTIPISALTREGVRD

LVQGFPGNLIIEHLPRGFSVADMTIDQIIHRRGVSAGVARSISITGLIPVING

EQLVDGGLLNMLPANVMCADTDCVITCDVLRFRVSKGGLPLPYTPPGILRL

TCGDNALPPIQETTLRAFDLAASTANIRELPRAVAILIEPDSKIGVINFKQIDALAE

GMMAARALQAQPDLYR"

9824..13294

/gene="MT2642"

9824..13294

/gene="MT2642"

REMARK Karl-von-Frisch-Strasse, Marburg 35032, Germany
 REFERENCE Sequence update by submitter
 5 (bases 1 to 35209)
 AUTHORS Mueller P. and Stingel D.
 TITLE Direct Submission
 JOURNAL Submitted (03-JAN-2000) Biology, Phillips Universitaet Marburg,
 Karl-von-Frisch-Strasse, Marburg 35032, Germany
 REMARK Sequence update by submitter
 COMMENT On Jan 3, 2000 this sequence version replaced gi:6136295.
 FEATURES
 location/Qualifiers
 1..35209

gene
 CDS
 /organism="Bradyrhizobium japonicum"
 /strain="USDA 110spc4"
 /db_xref="taxon:375"
 727..1647
 /gene="arsA"
 727..1647
 /note="similar to other arylsulfatases"
 /codon_start=1
 /product="putative arylsulfatase"
 /protein_id="AA04318.1"
 /db_xref="GI:6136297"

gene
 CDS
 /translation="MHGFDEFFGYLYHLDAMEDPAHRYNPOSILATVCGPRNVHSMAT
 WDDPTVOTRMKIGKOKIEDSGPLPKRMETVDEILKTFDIDOKARONKPPFVW
 LNPTRHVVYTHLSEKYENMRNSENGWSVEGGFALDDIVGAVMKIKIDECDDNTIV
 APTDGAENFTWPDGOTPPFAGSGKTAMEGFRPCMTBPGCKYPAGVENGISGL
 DMPTFVAAGNPNINGELQKGLGDKYKYVLOGYKPDGPDYKRPVRSPLDLFI
 GGTTERGAHRILOIPLHRSAPRLVGRHCGG"
 complement(2043..3872)
 /gene="sila"
 complement(2043..3872)
 /note="similar to Rickettsia prowazekii strain Madrid E"
 /codon_start=1
 /transl_table=11
 /product="putative soluble lytic transglycosylase
 precursor"

gene
 CDS
 /protein_id="AA04319.1"
 /db_xref="GI:6136298"
 /translation="MAPSDYRAFLSANPSMPSOFELRRLEAMMDRRDSDVAMSW
 FENESVAKGFTLAKAMLAGRDRANEPAGRLAORSYVCGHREORAPVRPAJ
 AGRSGADDTLSTAKTRPLRAKRLGTYALAKARLASVKAPNTRALLEAVRAC
 TWPVSGFLSEIOLLRREKFAEAQIMLSAPKDPNRLYLDPMWIERLARAKMID
 TTRAGTWGRARAEAAAGROQAEANAVARAEOSTSYGQLARAKLGIPOTELNSOPR
 LAGKRSYAAFPDFYAPVNGIPOTPIGPEVERSHRLRRPAGSERVOSIGLAGAG
 YGLMOTPDARVYCKRHGATYDLGRKNDASAVATYGSALGGLLEDYKOSTIMTA
 AYAGRGSVKRWVRYGDPDPKVDANDWVLLPFSETRNTAADHGEPSGLPRPLRK
 RHAVADGRPARPRRORIGITALLSIRREHRRAPDRRPA"

gene
 CDS
 /gene="dapa"
 4811..5701
 /note="involved in the biosynthesis of the rhizopine
 3-O-methyl-scyllo-inosamine"
 /function="catalyzes the first step in the biosynthesis of
 diaminopimelate and lysine from aspartate semialdehyde"
 /note="similar to the Mosa protein of Sinorhizobium
 meliloti; belongs to the DHDPS family"
 /codon_start=1
 /transl_table=11
 /product="dihydrodipicolinate synthase"
 /protein_id="AA04320.1"
 /db_xref="GI:6136299"

gene
 CDS
 /translation="MAAKTKRGSTFALVTPFKNSLDEAEFSLVNMOLISGJNGLY
 PVGTGESPTLSHDEHKVYEWCIIEAKGSRVAVAGNSNTRKEATELEAQAHEKAGAD
 AGLVAVYNNKPTOEGMIOHFKAINDAIGIPITITIPRSVITDMSVPMKRLMELKN
 IAGVDATASMYRVSQOARAGMEDNOLSGEAVTITIGVAGCGHGTSTVSNVAPRLC
 SEFTTAMOGDORFALKLHDKLMPVYNNILFIESNPATVYAMSLLGKLDLTLRLPMVP
 VPEPTRVAVRSAMVHAGLTIN"

gene
 CDS

5729..6202
 /gene="mscL"
 5729..6202
 /gene="mscL"
 /note="similar to large conductance mechanosensitive
 channel proteins of other microorganisms; putative
 mechanosensitive channel protein large subunit; two
 transmembrane domains"
 /codon_start=1
 /transl_table=11
 /product="mscL"
 /protein_id="AA04321.1"
 /db_xref="GI:6136300"

gene
 CDS

6236..6712
 /gene="smpB"
 6236..6712
 /note="small protein homolog"
 /codon_start=1
 /transl_table=11
 /product="SmpB"
 /protein_id="AA04322.1"
 /db_xref="GI:6136301"

gene
 CDS

6848..7297
 /gene="bcpB"
 6848..7297
 /note="bacterioferritin-comigrating protein"
 /codon_start=1
 /transl_table=11
 /product="BcpB"
 /protein_id="AA04323.1"
 /db_xref="GI:6136302"

CDS

translation="MLSALPGRTVFAVPRGPKISLVDDMDIPGARGTPOCA
 FRLIEFLKAGAAHVEGLSTOSNEYOTMASRLHLPPVLSDEKALARAKLPLME
 VAGIMLILKRALIIDDAKITKVPYVPEPPDRNAGVLDMLKMPKXD"
 complement(7306..8157)
 /note="ORF283; contains two putative transmembrane
 domains"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AA04324.1"
 /db_xref="GI:6136303"

CDS

translation="MADDECARPTSSSCNPSSAATRENAIRRAIVASARPATY
 PRNSSARRPRMFGATITRSEARSGQPPPLVVDRCPLCPRLVAREGESRARAV
 VAITPRSPGDKARILLVGLAPGOGANKRGPRFGDYAGDLLVATLLEFGPAKAY
 QARPDDGLKLVDCRIANVHCVPONKPLPAEINRCROFLANLATMNLRAIYALGR
 IADHVLKPLNLKSGQAPFGHGAFAHQAQAFRLYDSYHCSRYNTNTGVTLPKFRSVS
 KVKADLD"
 complement(8154..8603)
 /note="ORF149"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AA04325.1"
 /db_xref="GI:6136304"

gene
 CDS

translation="MSPTNKIALFTIDGAMLYATATKLGFDIDYKRLAEFOSRGL
 LRAEYVTAIIEDEYSIRPLIIMLDYNGTVVTKATKEFTDASGRKVGNDIELA
 VDAMFLAHIDQWVLESGDGFSLVAVORGVYVSRSPASRR"
 8929..9243
 /gene="rnpO"
 8929..9243
 /note="RNA-dependent nucleotide polymerase omega subunit
 homolog"
 /codon_start=1


```

CDS
10834..12516
/gene="PAB1"
/note="binds Poly(A) tail of mRNA; Biasip similarity to
many polyadenylate-binding proteins from several
organisms; contains RNA recognition motifs and bacterial
histidine kinase domain"
/codon_start=1
/product="PolyA Binding Protein 1"
/protein_id="AA070533.1"
/db_xref="GI:7882851"
/translation="MAAAGEAAPVAAHOPQMDKPMQIASIVYDGLDATTINPOLVEL
FKPPGTLINWVACDITLORSIGYVNFHDSAEKALIESMNFKEGKCVLTMQO
RDPALRYSGNGNVEVKILKEDVDKSLHDIETFKGSLTSGWODEGSRGCVFVH
KDETSAKDAIVKMGADHASDEKALVYANFTIRNRRLAALVANTYIKOVLPIV
NKDYIEKFEPAKFGGITSAACKDKSGRVFAFCNEKHDDAVKAEAMHDIIDGITAP
SCRWVSESGVSRGPGFVSFNADENAAALRENGRMILNGKPLIVNIAORDRYML
RLOPQORLOMMRMOMQMPMPVSGSGRPMGRGROOLGGAQGHMPMPSPQOPAP
AOPQGFATPSAVGVQATPKHSPGDVEETPLPITPOELESMSPOEORALGDRFL
KYVEIAPELAKITGMFLMKPKAEVELINDOKLEERTVETALCVLKAHQTA"
15582..17570
/gene="511.2"
/note="L3342.6; predicted using Glimmer, Testcode and
Codonusage, Glimmer predicts start codon at 15624"
15582..17570
/gene="511.2"
/codon_start=1
/product="511.2"
/protein_id="AA070534.1"
/db_xref="GI:7882852"
/translation="MTSNPYIAOHGSCMEVWPNNGHEAEDVDAALGERRPNAPMHTIT
VRRSKNLITAITTTAASSSALGVHNTYOTLTNPALANLRHGLRVDDPSPRPPOD
ANSATSLPPLPADRLSVGSDGLREKGETEVREWKAPNTNSTSPNGAEGDADA
AINGNAHGLEIVSTAPASRGTEGSGCTSAVPATVPATVPATVPATVPATVPATVP
SPESLIVRKSTASTFTPLQAEVSSAPTVDALYRLDVIOMAOVILDKOISGDA
RALPAPAMPPEPAESAOHEOQTAVLAAQTSRANEDVMPEDSLDCVPLASTHAA
LYALSAERGVGVPAASTPAKAVADVAVPDETRVAVCAVSPASIAIMERLRRK
GPATRAPMGLTSPAHSTFGCAETGAKRTISRKIAEVPSSAAVSATVAQULTAA
GOLORSSPNAGGEGHPSLRAPQSGSAGSVSRVHEDPSLIPGLAEFLRAGNPNSF
TSLAATIHAPPAKASPTRRPRRSGPPAADSPAVLGHSPSLORSSGTA
LSQADAVATPSLHLPIYAHTEGSAIVADVSPRGSARARRVSTPSYAOPTLSWLKSGS
EVSADQFPMRSRYASPOTSLKDPVTASRDAPSEPAL"
19700..21268
/gene="511.3"
/note="ORFalpha: L3342.7; predicted using Glimmer,
Testcode and Codonusage"
19700..21268
/gene="511.3"
/note="Prosite predicts bipartite nuclear localization
signal"
/codon_start=1
/product="511.3"
/protein_id="AA070535.1"
/db_xref="GI:7882853"
/translation="MLSDAEAKITVAVLKSLDPRRTTYEVENLYOLNCFRIAVRAV
LADMSGDVLDQCEKCGEAKOVAKRAAGVATAADNAPSLDFALPEAGGONAEI
ESATSAAHGSRKRSKSGEKKKKEKRRKRLDDEPLVAPILPAGGSGCEN
NAADKATSSPSTRWVLSRDDLTSSSEVLPALSTSDYKCAVANTPLARFTPEVT
RDGASCCOAVELYRCVVIAMRYGEPQLSMACEKMQSOVAASKAGTIEDGDSRS
AKRAGSRPKTSSDGEDEGADSDTSLEVENTLESLIAQYDOQAEEDEADITL
LPRIISERFOORLHGTDEWVILLCHGQYFAGGTCITHKAFORVVRKKGQ
KQSSNAKDAGSYNSVQIYAVQEVKRWVVDRLIDMMPYIOASPIIYAAPGPORR
AVLTFDSLPAVASVGRKGVSPLOLKDPVSRVPLTTHRPTEEVKRIYGVSRCSL
LYRQEEAOERNTN"
23224..23730
/gene="511.4"
/note="ORF2; L3342.8; predicted using Glimmer, Testcode
and Codonusage"
23224..23730
/gene="511.4"
/codon_start=1
/product="511.4"
/protein_id="AA070536.1"

```

```

Query Match
Best Local Similarity 100.0%; Pred No. 4.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 cgcagcagatgcac 16
Db 7912 CGCAGCAGGTGCATC 7898

```

Search completed: December 26, 2001, 12:51:46
Job time: 8828 sec

```

/db_xref="GI:7882854"
/translation="MSDYSDDEDVVLQVCANRRHCOGIDLEFDEETNOSYCGRCRA
LVARTESEFRVLLNSDDLGIVKILFDQFEDNSRGFMVYKEMNAFOEYTERGSDPPIE
SSAALKEFPEKEEYDIDISAHGKDGAIVLLVDLENMYGGYLVNNVDALVEDEANMYGQ

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:16:40 ; Search time 368.09 Seconds
(without alignments)
58.228 Million cell updates/sec

Title: US-09-396-196f-5
Perfect score: 25
Sequence: 1 gcgcagcagtgatcgcgcagcatt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428652619 residues

Word size : 9

Total number of hits satisfying chosen parameters: 46648

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

N.Geneseq_1101:*

- 1: /SID2/gcgcdata/geneseq/geneseq/NA1980.DAT:*
- 2: /SID2/gcgcdata/geneseq/geneseq/NA1981.DAT:*
- 3: /SID2/gcgcdata/geneseq/geneseq/NA1982.DAT:*
- 4: /SID2/gcgcdata/geneseq/geneseq/NA1983.DAT:*
- 5: /SID2/gcgcdata/geneseq/geneseq/NA1984.DAT:*
- 6: /SID2/gcgcdata/geneseq/geneseq/NA1985.DAT:*
- 7: /SID2/gcgcdata/geneseq/geneseq/NA1986.DAT:*
- 8: /SID2/gcgcdata/geneseq/geneseq/NA1987.DAT:*
- 9: /SID2/gcgcdata/geneseq/geneseq/NA1988.DAT:*
- 10: /SID2/gcgcdata/geneseq/geneseq/NA1989.DAT:*
- 11: /SID2/gcgcdata/geneseq/geneseq/NA1990.DAT:*
- 12: /SID2/gcgcdata/geneseq/geneseq/NA1991.DAT:*
- 13: /SID2/gcgcdata/geneseq/geneseq/NA1992.DAT:*
- 14: /SID2/gcgcdata/geneseq/geneseq/NA1993.DAT:*
- 15: /SID2/gcgcdata/geneseq/geneseq/NA1994.DAT:*
- 16: /SID2/gcgcdata/geneseq/geneseq/NA1995.DAT:*
- 17: /SID2/gcgcdata/geneseq/geneseq/NA1996.DAT:*
- 18: /SID2/gcgcdata/geneseq/geneseq/NA1997.DAT:*
- 19: /SID2/gcgcdata/geneseq/geneseq/NA1998.DAT:*
- 20: /SID2/gcgcdata/geneseq/geneseq/NA1999.DAT:*
- 21: /SID2/gcgcdata/geneseq/geneseq/NA2000.DAT:*
- 22: /SID2/gcgcdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	839	22	AA162941 Human genomic DNA
2	25	100.0	1041	20	AA01303 E. coli biotin syn
3	25	100.0	1084	10	AA01329 E. coli Bio B gene
4	25	100.0	1121	7	AA060496 Sequence encoding
5	25	100.0	5872	15	AA062386 Biotin-biosynthesis
6	16	64.0	1498	21	AA051224 Arabidopsis thaliana
7	16	64.0	1500	21	AA036258 Arabidopsis thaliana
8	16	64.0	4674	22	AA018657 Human cDNA sequence
9	15	60.0	632	21	AA011615 Aspergillus niger
10	15	60.0	2597	21	AA033313 Human secreted pro
11	14	56.0	300	21	AA01491 Human colon cancer

C	12	14	56.0	360	22	AA137141	Probe #5827 used t
C	13	14	56.0	597	22	AA084601	E. coli growth and
C	14	14	56.0	667	22	AA071326	Corynebacterium gl
C	15	14	56.0	668	22	AA0502190	DNA encoding probe
C	16	14	56.0	674	22	AA07252	Human cDNA clone (
C	17	14	56.0	702	22	AA05704	C glutamicum codin
C	18	14	56.0	719	19	AA020504	Conus geographus c
C	19	14	56.0	719	19	AA017130	Conus geographus c
C	20	14	56.0	844	22	AA06472	Human cDNA clone (
C	21	14	56.0	943	22	AA048556	Human fascin DNA f
C	22	14	56.0	1116	22	AA064755	Human secreted pro
C	23	14	56.0	1131	21	AA06288	C glutamicum codin
C	24	14	56.0	1190	21	AA068831	Human pancreatic c
C	25	14	56.0	1240	22	AA048555	Human fascin DNA f
C	26	14	56.0	1278	22	AA06089	C glutamicum codin
C	27	14	56.0	1326	22	AA068314	C glutamicum codin
C	28	14	56.0	1344	21	AA056381	Escherichia coli f
C	29	14	56.0	1431	22	AA158241	Human polynucleotl
C	30	14	56.0	1449	22	AA048554	Human fascin DNA f
C	31	14	56.0	1619	22	AA048553	Human fascin DNA f
C	32	14	56.0	1851	22	AA017986	Human cDNA sequenc
C	33	14	56.0	1934	22	AA048552	Human fascin DNA f
C	34	14	56.0	2017	21	AA069762	Human ovarian carc
C	35	14	56.0	2193	17	AA030684	Kaposi's sarcoma a
C	36	14	56.0	2193	17	AA016814	Kaposi's sarcoma a
C	37	14	56.0	2193	20	AA020914	Nucleotide sequenc
C	38	14	56.0	2265	22	AA065400	C glutamicum codin
C	39	14	56.0	2397	22	AA061014	P. putida KT240-a
C	40	14	56.0	2497	22	AA014473	Human cDNA sequenc
C	41	14	56.0	3069	22	AA048551	Human fascin DNA f
C	42	14	56.0	3593	21	AA039011	Mouse Ees2 coding
C	43	14	56.0	4100	22	AA048581	Human fascin DNA f
C	44	14	56.0	4182	16	AA081778	DNA encoding cytd
C	45	14	56.0	4625	21	AA029010	Mouse Ees2 full le
C	46	14	56.0	4975	21	AA029027	Mouse Ees2L coding
C	47	14	56.0	5362	18	AA097296	Mouse additional s
C	48	14	56.0	5362	18	AA097294	Mouse additional s
C	49	14	56.0	5676	19	AA021186	Amycolatopsis medi
C	50	14	56.0	6014	21	AA029026	Mouse Ees2L cDNA s
C	51	14	56.0	12893	21	AA081733	N. meningitidis pa
C	52	14	56.0	13053	22	AA048621	Human fascin DNA f
C	53	14	56.0	15951	22	AA048620	Human fascin DNA f
C	54	14	56.0	16951	22	AA048622	Human fascin DNA f
C	55	14	56.0	20387	19	AA062159	HSV-2 strain SB5 C
C	56	14	56.0	20710	17	AA030681	Kaposi's sarcoma a
C	57	14	56.0	20710	17	AA016806	Kaposi's sarcoma a
C	58	14	56.0	2638	19	AA062134	HSV-2 strain SB5 C
C	59	14	56.0	35100	20	AA073803	KSHV LTR DNA (nucl
C	60	14	56.0	53789	19	AA021187	Amycolatopsis medi
C	61	14	56.0	117213	19	AA062176	HSV-2 strain SB5 C
C	62	14	56.0	137507	19	AA019941	KSHV long unique c
C	63	14	56.0	309400	22	AA068534	C glutamicum codin
C	64	14	56.0	349980	22	AA068534	Neisseria meningit
C	65	14	56.0	349980	22	AA068525	C glutamicum codin
C	66	14	56.0	349980	22	AA068526	C glutamicum codin
C	67	14	56.0	349980	22	AA068527	C glutamicum codin
C	68	14	56.0	349980	22	AA068528	C glutamicum codin
C	69	14	56.0	21	14	AA041769	Myc EMSA probe pro
C	70	13	52.0	21	20	AA033237	wheat viviparous 1
C	71	13	52.0	22	19	AA011150	Oligonucleotide #3
C	72	13	52.0	51	22	AA089605	Human DNA/RNA bind
C	73	13	52.0	142	22	AA123228	Probe #13161 for g
C	74	13	52.0	142	22	AA148545	Probe #17231 used
C	75	13	52.0	142	22	AA108872	Probe #863 used t
C	76	13	52.0	286	21	AA045502	Human secreted exp
C	77	13	52.0	300	20	AA021474	Human gene express
C	78	13	52.0	300	20	AA021476	Human gene express
C	79	13	52.0	300	20	AA021298	Human gene express
C	80	13	52.0	300	21	AA021298	Human gene express
C	81	13	52.0	336	21	AA005494	Human secreted pro
C	82	13	52.0	336	22	AA014012	Probe #3945 for ge
C	83	13	52.0	336	22	AA013589	Probe #4075 used t
C	84	13	52.0	396	22	AA0105870	Probe #3861 used t

```

C 85      13      52.0      396      22      AAF64927      Novel human polyu
C 86      13      52.0      412      22      AAF64902      Novel human polyu
C 87      13      52.0      413      21      AAC02052      Human secreted pro
C 88      13      52.0      414      16      AAO87078      Halomonas KS-3 str
C 89      13      52.0      447      17      AAT27962      Hepatitis C virus
C 90      13      52.0      471      22      AAI59982      Human polynucleoti
C 91      13      52.0      538      20      AAZ40788      Secreted protein F
C 92      13      52.0      549      21      AAC93854      Cat flea hindgut a
C 93      13      52.0      571      18      AAT69744      Corn starch branch
C 94      13      52.0      629      21      AAF09259      Fusarium venenatum
C 95      13      52.0      632      21      AAF09268      Fusarium venenatum
C 96      13      52.0      636      22      AAS07130      DNA encoding Group
C 97      13      52.0      646      21      AAC55942      Eucalyptus grandis
C 98      13      52.0      689      22      AAH98791      Human EST-derived
C 99      13      52.0      725      22      AAF22436      Human breast cance
C 100     13      52.0      749      20      AAZ16470      Human gene express

```

ALIGNMENTS

```

RESULT 1
AAI62941
ID      AAI62941 standard; DNA; 839 BP.
XX
AC      AAI62941;
XX
DT      22-OCT-2001 (first entry)
XX
DE      Human genomic DNA SEQ ID NO 269.
XX
KW      Human; nocotropic; neuroprotective; cytostatic; dermatological; virucide;
KW      immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW      antiparkinsonian; antistickling; antianaemic; antiarthritic; cancer;
KW      antitubercular; hepatocytropic; cerebroprotective; antiinflammatory;
KW      antiallergic; antidiabetic; antilucer; anticomvulsant; antifungal;
KW      antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW      neurological disease; infection; nephrotropic; gene therapy; vaccine;
KW      ds.
XX
OS      Homo sapiens.
XX
PN      WO20015449-A1.
XX
PD      02-AUG-2001.
XX
PF      17-JAN-2001; 2001WO-US01346.
XX
PR      31-JAN-2000; 2000US-0179065.
PR      04-FEB-2000; 2000US-0180628.
PR      19-MAY-2000; 2000US-0205515.
PR      07-JUL-2000; 2000US-0216880.
PR      14-JUL-2000; 2000US-0218290.
PR      14-AUG-2000; 2000US-0225447.
PR      01-SEP-2000; 2000US-0229343.
PR      06-SEP-2000; 2000US-0230437.
PR      08-SEP-2000; 2000US-0231243.
PR      25-SEP-2000; 2000US-0234997.
PR      29-SEP-2000; 2000US-0236367.
PR      13-OCT-2000; 2000US-0239937.
PR      08-NOV-2000; 2000US-0246476.
PR      08-NOV-2000; 2000US-0246477.
PR      08-NOV-2000; 2000US-0246525.
PR      08-NOV-2000; 2000US-0246526.
PR      08-NOV-2000; 2000US-0246528.
PR      17-NOV-2000; 2000US-0249210.
PR      17-NOV-2000; 2000US-0249211.
PR      17-NOV-2000; 2000US-0249214.
PR      17-NOV-2000; 2000US-0249265.
PR      01-DEC-2000; 2000US-0250160.
PR      01-DEC-2000; 2000US-0250391.
PR      05-DEC-2000; 2000US-0251030.
PR      05-DEC-2000; 2000US-0251988.

```

```

PR      05-DEC-2000; 2000US-0256719.
PR      06-DEC-2000; 2000US-0251479.
PR      08-DEC-2000; 2000US-0251989.
PR      08-DEC-2000; 2000US-0251990.
PR      11-DEC-2000; 2000US-0254097.
XX
XX      (HUMA-) HUMAN GENOME SCI INC.
XX
XX      Rosen CA, Barash SC, Ruben SM.
XX
XX      WPI; 2001-476225/51.
XX
XX      Novel plasma membrane associated proteins useful for diagnosing,
XX      treating, preventing and/or prognosing disorders related to the
XX      proteins, including cancer, immune response and neuronal disorders -
XX
XX      Example 2; SEQ ID NO 269; 532pp + Sequence Listing; English.
XX
CC      The invention relates to novel genes (AAI62752-AAI62961) and proteins
CC      (AAM42347-AAM42415) useful for preventing, treating or ameliorating
CC      medical conditions e.g. by protein or gene therapy. The genes are
CC      isolated from a range of human tissues disclosed in the specification.
CC      The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC      in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC      and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC      marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC      (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC      haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC      disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC      colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC      (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC      epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC      and parasitic infections.
CC      Note: The sequence data for this patent did not form part of the
CC      printed specification, but was obtained in electronic format directly
CC      from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX      Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other;
XX

```

```

Query Match      100.0%; Score 25; DB 22; Length 839;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 ggcgcagcaggtgcagtcgcagcatt 25
      |||
DB      111 ggcgcagcaggtgcagtcgcagcatt 135

```

```

RESULT 2
AAAX01303
ID      AAAX01303 standard; DNA; 1041 BP.
XX
AC      AAAX01303;
XX
DT      12-APR-1999 (first entry)
XX
DE      E. coli biotin synthetase (BioB) coding sequence.
XX
KW      DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW      biotin synthase; biotin production; vitamin H; BioB; ss.
XX
OS      Escherichia coli.
XX
XX      US5869719-A.
XX
XX      09-FEB-1999.
XX
XX      30-APR-1997; 97US-0846338.
XX
XX      30-APR-1997; 97US-0846338.
XX
XX      08-MAR-1995; 95US-0401068.
XX

```

PA (NOVS) NOVARTIS FINANCE CORP.
 XX
 PI Patton DA:
 XX
 DR WPI: 1999-152902/13.
 DR P-PSDB; AAN73906.
 XX
 PT Transgenic plants with high biotin levels - transformed with DNA
 PT encoding di- amino-pelargonic acid amino-transferase or biotin
 PT synthase
 XX
 PS Example 2; Column 37-40; 34pp; English.
 CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC di- amino-pelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.
 CC
 XX
 SQ Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;
 Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.00023;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 ggcgcagcagtgatcgcgcagcatt 25
 Db 79 ggcgcagcagtgatcgcgcagcatt 103
 RESULT 3
 AAN91329
 ID AAN91329 standard; DNA: 1084 BP.
 XX
 AC AAN91329;
 XX
 DT 15-FEB-1990 (first entry)
 DE E.coli Bio B gene.
 XX
 KW E.coli Bio B gene; biotin.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT CDS 24..1064
 FT /*tag=a
 XX
 PN GB2216530-A.
 PD 11-OCT-1989.
 XX
 PF 17-MAR-1989; 89GB-0006210.
 XX
 PR 22-MAR-1988; 88GB-0006804.
 PR 17-MAR-1989; 89GB-0006210.
 XX
 PA (UKAG-) UK MTN. AGRIC. FISH.
 XX
 PI Pearson BM, McKee RA;
 XX
 DR WPI: 1989-295085/41. P-PSDB P91392
 XX
 PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.
 XX
 PS Table 3; page 33-4; 52pp; English.
 XX
 CC The gene can be used in a plasmid for expression of enzymes of the biotin

CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
 CC Lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.
 CC
 XX
 SQ Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;
 Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 0.00023;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 ggcgcagcagtgatcgcgcagcatt 25
 Db 102 ggcgcagcagtgatcgcgcagcatt 126
 RESULT 4
 AAN60496
 ID AAN60496 standard; DNA: 1121 BP.
 XX
 AC AAN60496;
 XX
 DT 17-OCT-1991 (first entry)
 DE Sequence encoding biotin synthesising enzyme.
 XX
 KW Biotin synthetic enzyme; E.coli; desbiobiotin; ds.
 XX
 FH Key Location/Qualifiers
 FT CDS 42..1082
 FT /*tag=a
 XX
 PN JP61149091-A.
 PD 07-JUL-1986.
 XX
 PF 24-DEC-1984; 84UP-0272605.
 XX
 PR 24-DEC-1984; 84UP-0272605.
 XX
 PA (NIPS) NIPPON SODA KK.
 XX
 DR WPI: 1986-216622/33.
 DR P-PSDB; AAP60536.
 XX
 PT Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
 XX
 PS Disclosure; Page 534; 23pp; Japanese.
 XX
 CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing desbiobiotin.
 CC
 SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;
 Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 0.00023;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 ggcgcagcagtgatcgcgcagcatt 25
 Db 120 ggcgcagcagtgatcgcgcagcatt 144
 RESULT 5
 AAO62386
 ID AAO62386 standard; DNA: 5872 BP.
 XX

XX 18-OCT-2000 (first entry)
DT
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 67727.
DE
XX Hybridisation assay: genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
OS
PN EPI033405-A2.
XX
XX 06-SEP-2000.
PD
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139493.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145513.
PR 27-JUL-1999; 99US-0145518.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148317.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.

```
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158363.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161982.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 64.0%; Score 16; DB 21; Length 1498;
Best Local Similarity 100.0%; Pred. No. 10;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 aggtgcacgcgcagca 23
Db 84 aggtgcacgcgcagca 99

RESULT 7
AAC36258
ID AAC36258 standard; DNA: 1500 BP.
XX
AC AAC36258;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 13129.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
protein identification; signal transduction pathway;

```
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
FN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
FE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134491.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
```


PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150366.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.

PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 64.0%; Score 16; DB 21; Length 1500;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 aggtgcatgcgcagca 23
|||||
Db 84 aggtgcatgcgcagca 99

RESULT 8
AAH18667/c
ID AAH18667 standard; cDNA: 4674 BP.
AC AAH18667;
XX
XX 26-JUN-2001 (first entry)
DT
XX Human cDNA sequence SEQ ID NO:18909.
DE
XX
XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
KW
XX Homo sapiens.
OS
XX EPI074617-A2.
PN
XX 07-FEB-2001.
PD
XX
PF 28-JUL-2000; 2000EP-0116126.

```

XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 8; SEQ ID 18909; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination
CC of the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AA892446 to
CC AA895883 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 4674 BP; 991 A; 1342 C; 1239 G; 1102 T; 0 other;
SO

```

Query Match 64.0%; Score 16; DB 22; Length 4674;
 Best Local Similarity 100.0%; Pred. No. 9.8;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 9 ggtgcatgcacagat 24
DB 2446 GTGTCATGCCAGCAT 2431

```

RESULT 9
 ID AAF11615/c
 ID AAF11615 standard; cDNA: 612 BP.
 AC AAF11615;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Aspergillus niger EST SEQ ID NO:4138.
 XX
 XX Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Aspergillus niger.

```

XX PN W0200056762-A2.
XX
XX PD 28-SEP-2000.
XX
XX PF 22-MAR-2000; 2000WO-US07781.
XX
XX PR 22-MAR-1999; 99US-0273623.
XX
XX PA (NOVO) NOVO NORDISK BIOTECH INC.
XX PA (NOVO) NOVO NORDISK AS.
XX
XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
PI WPI: 2000-594572/56.
XX
XX Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
XX Claim 87; Page 1813; 3161pp; English.
XX
XX The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
XX Sequence 612 BP; 147 A; 162 C; 144 G; 156 T; 3 other;
SO

```

Query Match 60.0%; Score 15; DB 21; Length 612;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 3 gcacgagtgatcagc 17
DB 118 CCAGCAGGTGCATCC 104

```

RESULT 10
 ID AAC93313
 ID AAC93313 standard; cDNA: 2597 BP.
 AC AAC93313;
 XX
 DT 16-FEB-2001 (first entry)
 XX
 DE Human secreted protein cDNA sequence #4.
 XX
 XX Cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;
 KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;

KW neurological disease; infection; human; secreted protein; ss.
 XX Homo sapiens.
 OS
 XX WO200058495-A1.
 PN
 XX
 PD 05-OCT-2000.
 XX
 XX 23-MAR-2000: 2000MO-US07661.
 PE
 XX 26-MAR-1999: 99US-0126504.
 PR
 PR 07-JAN-2000: 2000US-0174847.
 XX
 PA (HOMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Ruben SM, Komatsoulis G;
 XX
 XX WPI: 2000-611720/58.
 DR
 DR P-PSDB: AAB51383.
 XX
 PT New nucleic acid molecules encoding 45 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 PS
 PS Claim 1: Pages 327-328; 410pp; English.
 XX
 XX The invention relate to the isolation of genes AAC93310-C93354 encoding
 CC 45 human secreted proteins AAB51380-B51423. The genes can be used to
 CC generate fusion proteins by linking to the gene for the human
 CC immunoglobulin G Fc portion (AAC93301) for increasing the stability of
 CC the fusion protein as compared to the human protein only. The genes and
 CC proteins are useful for preventing, ameliorating or treating medical
 CC conditions, e.g. by protein or gene therapy. The genes are isolated
 CC from a range of human tissues disclosed in the specification. The
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d)
 CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 XX
 XX Sequence 2597 BP: 721 A; 604 C; 534 G; 737 T; 1 other;
 SO

Query Match 60.0%; Score 15; DB 21; Length 2597;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 gcagcagtcacgccca 20
 ||||||||||||||||
 Db 2256 gcagcagtcacgccca 2270

RESULT 11
 AAA01491/c
 ID AAA01491 standard; cDNA; 300 BP.
 XX
 AC AAA01491;
 XX
 DT 19-MAY-2000 (first entry)
 XX
 DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:1482.
 XX
 KW Human; colon cancer; tumour; diagnosis; gene expression product;
 KW probe; detection; cancerous state; metastasis; identification;
 KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
 KW oestrogen receptor-negative breast cancer; lung cancer; ss.
 XX

OS Homo sapiens.
 XX
 XX WO9958675-A2.
 PN
 XX
 PD 18-NOV-1999.
 XX
 XX 13-MAY-1999: 99MO-US10602.
 PE
 XX 14-MAY-1998: 98US-0085426.
 PR
 PR 15-MAY-1998: 98US-0085537.
 PR 15-MAY-1998: 98US-0085696.
 PR 21-OCT-1998: 98US-0105234.
 PR 27-OCT-1998: 98US-0105877.
 XX
 XX (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEO INC.
 XX
 PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
 PI Reinhard C, Glese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
 PI Lamson G, Drmanac R, Ckrvenjakov R, Dickson M, Drmanac S, Labat I;
 PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Grain B;
 XX
 XX WPI: 2000-126369/11.
 DR
 XX Polynucleotide library used to determine cancerous states of mammalian
 PT cells -
 PT
 PS Claim 1: Page 536; 1097pp; English.
 XX
 XX AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
 CC libraries constructed from human colon cancer cell lines. The present
 CC invention also describes a method of detecting differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell, comprising
 CC detecting at least one differentially expressed gene product in a test
 CC sample derived from a cell suspected of being cancerous, where detection
 CC of the differentially expressed gene product is correlated with a
 CC cancerous state of the cell from which the test sample was derived.
 CC The polynucleotide sequences can be used in a method for detecting
 CC differentially expressed genes correlated with a cancerous state of a
 CC mammalian cell. The polynucleotides can also be used as probes for
 CC detecting and mapping related genes. They can be used in diagnosis and
 CC prognosis of diseases and disorders (e.g. identification of
 CC pre-metastatic or metastatic cancerous states, stages of cancer, or
 CC responsiveness of cancer to therapy). This is particularly for breast
 CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
 CC negative breast cancer, lung cancer, and colon cancer.
 XX
 XX Sequence 300 BP: 56 A; 89 C; 94 G; 61 T; 0 other;
 SO

Query Match 56.0%; Score 14; DB 21; Length 300;
 Best Local Similarity 100.0%; Pred. No. 11e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcagcagtcacalc 16
 ||||||||||||||||
 Db 108 GCAGCAGTCGCATC 95

RESULT 12
 AA137141/c
 ID AA137141 standard; DNA; 360 BP.
 XX
 AC AA137141;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Probe #5827 used to measure gene expression in human placenta sample.
 XX
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.
 XX
 OS Homo sapiens.

XX
 PN WO200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI: 2001-488897/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 PS Claim 25: SEQ ID No 5837; 654pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX
 SQ Sequence 360 BP; 106 A; 87 C; 69 G; 98 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 360;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 gcagcaggtgcatc 16
 |||||
 DB 218 GCAGCAGGTGCATC 205

RESULT 13

AAH84601/C
 ID AAH84601 standard; DNA; 597 BP.

XX
 AC AAH84601;

DT 26-SEP-2001 (first entry)

XX E. coli growth and proliferation related coding sequence SEQ ID NO:229.

XX Escherichia coli; growth; proliferation; microbial; antimicrobial;

KW bacterial infection; microorganism; ds.

XX Escherichia coli.

OS WO200134810-A2.

XX
 PN 17-MAY-2001.

XX
 PD 09-NOV-2000; 2000WO-US30950.

XX
 PF 09-NOV-1999; 99US-0164415.

XX
 PR (ELIT-) ELITRA PHARM INC.

XX
 PI Forsyth RA, Ohlsen K, Zyskind J;

XX
 DR WPI: 2001-335933/35.

DR
 P-PSDB; AAG98930.

XX
 PT Novel nucleic acids that inhibit Escherichia coli proliferation, useful
 PT for screening for homologous genes and for designing expression vectors
 XX
 PS Claim 9; Page 275-276; 522pp; English.

XX
 CC AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
 CC related DNA sequences (1). AAH84500 to AAH84670 encode the E. coli
 CC growth and proliferation related proteins given in AAG99078 and AAG98830
 CC to AAG98999. (1) can be used as potential targets for the generation of
 CC new antimicrobial agents, and for identification of compounds which
 CC interact with the gene products of (1). In addition the expression of
 CC (1) and the purification of the proteins, the purified proteins can be
 CC used to generate reagents and screen small molecule libraries or other
 CC candidate compound libraries for compounds that can be further developed
 CC to yield novel antimicrobial compounds. In addition, nucleic acid probes
 CC complementary to (1) that are specific for particular species of
 CC microorganisms can be used to identify particular microorganism species
 CC in clinical specimens, therefore, providing a rapid and dependable
 CC method by which to identify the causative agents of a bacterial
 CC infection. Also, antibodies generated against proteins translated from
 CC mRNA transcribed from proliferation-regulated sequences can also be used
 CC to screen for specific microorganisms that produce such proteins in a
 CC species-specific manner. AAH84371 and AAH84670 represent sequencing
 CC primers used in the isolation of E. coli growth and proliferation
 CC related sequence, which are used in an example from the present
 CC invention.
 XX
 SQ Sequence 597 BP; 130 A; 158 C; 184 G; 125 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 597;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 tgcacgcacgat 24
 |||||
 DB 275 TGCATGCCACGAT 262

RESULT 14

AAE71326
 ID AAE71326 standard; DNA; 667 BP.

XX
 AC AAE71326;

DT 30-APR-2001 (first entry)

XX Corynebacterium glutamicum HA protein nucleotide sequence SEQ ID NO:377.

XX Corynebacterium glutamicum; homeostasis; adaptation; HA protein;

KW fine chemical production; organic acid; pyrimidine base; nucleoside;

KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;

KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;

KW carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;

KW diagnosis; Corynebacterium diptheriae; genetic engineering;

KW Brevibacterium; environmental condition; ds.

XX
 OS Corynebacterium glutamicum.

XX
 PN WO200100842-A2.

XX
 PD 04-JAN-2001.

XX
 PF 23-JUN-2000; 2000WO-IB00911.

XX
 PR 25-JUN-1999; 99US-0141031.

XX
 PR 08-JUL-1999; 99DE-1031636.

XX
 PR 09-JUL-1999; 99DE-1032126.

XX
 PR 09-JUL-1999; 99DE-1032126.

XX
 PR 09-JUL-1999; 99DE-1032127.

XX
 PR 09-JUL-1999; 99DE-1032128.

PR 09-JUL-1999; 99DE-1032129.
 PR 09-JUL-1999; 99DE-1032226.
 PR 14-JUL-1999; 99DE-1032920.
 PR 14-JUL-1999; 99DE-1032922.
 PR 14-JUL-1999; 99DE-1032924.
 PR 14-JUL-1999; 99DE-1032928.
 PR 14-JUL-1999; 99DE-1032930.
 PR 14-JUL-1999; 99DE-1032933.
 PR 14-JUL-1999; 99DE-1032935.
 PR 14-JUL-1999; 99DE-1032973.
 PR 14-JUL-1999; 99DE-1033002.
 PR 14-JUL-1999; 99DE-1033003.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1033006.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041390.
 PR 31-AUG-1999; 99DE-1041391.
 PR 03-SEP-1999; 99DE-1042088.
 XX
 PA (BADI) BASF AG.
 PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;
 DR WPI: 2001-061974/07.
 DR P-PSDB: AAB79211.
 XX
 PT New isolated Corynebacterium glutamicum nucleic acid for production or
 PT modulation of production of fine chemicals such as amino acids,
 PT nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins
 PT or enzymes -
 XX
 PS Claim 3; Page 630-631; 712pp; English.
 XX
 CC AAF7138 to AAF71357 encode the Corynebacterium glutamicum homeostasis
 CC and adaptation (HA) proteins given in AAB79023 to AAB79242. The
 CC C. glutamicum HA genes (I) can be used in vectors for expression in host
 CC cells and production of fine chemicals, such as, an organic acid,
 CC proteogenic or nonproteogenic amino acid (preferred), purine or
 CC pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated
 CC fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,
 CC polyketide or enzyme. The amino acids produced can be lysine, glutamine,
 CC glutamate, alanine, aspartate, glycine, serine, threonine, methionine,
 CC cysteine, valine, leucine, isoleucine, arginine, proline, histidine,
 CC tyrosine, phenylalanine, or tryptophan. The fine chemical production can
 CC be modulated. The presence of (I) or HA proteins encoded by then are
 CC used for diagnosing the presence or activity of Corynebacterium
 CC diptheriae. (I) can be used to map the C. glutamicum genome or can be
 CC used as markers for genetically engineered Corynebacterium or
 CC Brevibacterium. The HA proteins encoded by the (I) are used to maintain
 CC homeostasis in C. glutamicum or help the microorganism to adapt to
 CC different environmental conditions.
 XX
 SQ Sequence 667 BP; 146 A; 210 C; 172 G; 139 T; 0 other:

Query Match 56.0%; Score 14; DB 22; Length 667;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcgcagcaggttca 14
 |||||
 DB 168 gcgcagcaggttca 181

RESULT 15
 AAS02190/C
 ID AAS02190 standard: cDNA: 668 BP.
 XX
 AC AAS02190;
 XX
 DT 29-AUG-2001 (first entry)
 XX

DE DNA encoding propeptide of conopeptide L5V, K15A.
 XX
 KW Gamma carboxyglutamate: conopeptide; neurological disorder; epilepsy;
 KW hypoxia; anoxia; ischaemia; stroke; brain; spinal cord; suffocation;
 KW myocardial infarct; trauma; drowning; perinatal asphyxia; hypoglycaemia;
 KW myodegeneration; Alzheimer's disease; Huntington's disease;
 KW senile dementia; Amyotrophic lateral sclerosis; multiple sclerosis;
 KW Parkinson's disease; Down's Syndrome; Korsakoff's disease; schizophrenia;
 KW AIDS; acquired immunodeficiency syndrome; HIV; neuronal damage; pain;
 KW seizure; chemical toxicity; addiction; dystonia; psychiatric disorder;
 KW mood disorder; memory; ophthalmic; parasitic worm; ss.
 XX
 OS Conus geographus.
 XX
 FH Key Location/Qualifiers
 FT 93..395
 FT CDS /ftag= a
 FT /product= "Propeptide of Conopeptide L5V, K15A"
 XX
 PN W0200118033-A1.
 XX
 PD 15-MAR-2001.
 XX
 PF 08-SEP-2000; 2000MO-US24816.
 XX
 PR 10-SEP-1999; 99US-0153034.
 PR 21-JUL-2000; 2000US-0219673.
 XX
 PA (UTAH) UNIV UTAH RES FOUND.
 PA (COGN) COGNETIX INC.
 XX
 PI Olivera BM, McIntosh JM, Garrett JE, Walker CS, Watkins M;
 PI Jones RM;
 XX
 DR WPI: 2001-27379/28.
 DR P-PSDB: AA001503.
 XX
 PT New isolated gamma-carboxyglutamate containing peptide for treating or
 PT preventing neurological and psychiatric disorders e.g. epilepsy,
 PT Alzheimer's disease, migraine, chemical toxicity, dystonia, anxiety,
 PT and depression -
 XX
 PF Claim 4; Page 30-31; 102pp; English.
 XX
 PS The sequence represents the coding sequence of the propeptide of gamma
 CC carboxyglutamate-containing conopeptide L5V, K15A. The conopeptide is used
 CC for treating or preventing disorders in which the pathophysiology
 CC involves excess excitation of nerve cells by excitatory amino acids or
 CC agonists of heterogeneous ionotropic glutamate receptors or heterogenous
 CC G protein coupled glutamate receptors. The disorders may be neurological
 CC disorders, such as: (i) seizure associated with epilepsy; (ii) a
 CC neurotoxic injury associated with hypoxia, anoxia, ischaemia, stroke,
 CC cerebrovascular accident, brain or spinal cord trauma, myocardial
 CC infarct, physical trauma, drownings, suffocation, perinatal asphyxia, or
 CC hypoglycaemic events; (iii) neurodegeneration associated with Alzheimer's
 CC disease, Huntington's disease, senile dementia, Amyotrophic lateral
 CC sclerosis, multiple sclerosis, Parkinson's disease, Down's Syndrome,
 CC Korsakoff's disease, schizophrenia, AIDS (acquired immunodeficiency
 CC syndrome) dementia from HIV infection, HIV infection, multi-infarct
 CC dementia, Binswanger dementia and neuronal damage associated with
 CC uncontrolled seizures; (iv) pain which is a migraine, acute pain, or
 CC persistent pain; (v) chemical toxicity which is addiction, morphine,
 CC opiate, oploid and barbiturate tolerance; and (vi) dystonia, urinary
 CC incontinence, muscle relaxation or sleep disorder. The disorders may be
 CC psychiatric disorders, such as, anxiety, major depression, manic-
 CC depressive illness, obsessive-compulsive disorder, schizophrenia, or mood
 CC disorders (bipolar disorder, unipolar depression, dysthymia, or seasonal
 CC effective disorder). The conopeptide is also used to treat memory or
 CC cognitive deficits, ophthalmic indications, or to control nematodes or
 CC parasitic worms.
 XX
 SQ Sequence 668 BP; 187 A; 160 C; 145 G; 176 T; 0 other:

Query Match 56.0%; Score 14; DB 22; Length 668;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 caggtcgcgcga 20
 |||
 Db 101 CAGGTGCATCGCCA 88

RESULT 16

AAH07252/C
 ID AAH07252 standard; CDNA; 674 BP.

AC AAH07252;

DT 26-JUN-2001 (first entry)

DE Human CDNA clone (5'-primer) SEQ ID NO:4087.

KW Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

PS Claim 1; SEQ ID 4087; 2537bp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
 CC full-length CDNs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length CDNs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length CDNs. The primers allow obtaining of the full-length
 CC CDNs easily without any special methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human CDNA sequences; AAB92446 to
 CC AAB95693 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX

SQ Sequence 674 BP; 169 A; 152 C; 180 G; 167 T; 6 other;

Query Match 56.0%; Score 14; DB 22; Length 674;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 gcagcaggtgcac 16
 |||
 Db 390 GCAGCAGGTGCATC 377

RESULT 17

AAH65704
 ID AAH65704 standard; DNA; 702 BP.

AC AAH65704;

DT 26-SEP-2001 (first entry)

DE C glutamicum coding sequence fragment SEQ ID NO: 739.

KW Corynebacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis; ds.

OS Corynebacterium glutamicum.

PN EP108790-A2.

PD 20-JUN-2001.

PF 18-DEC-2000; 2000EP-0127688.

PR 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOGYO KK.

PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

DR WPI; 2001-376931/40.

PS Claim 8; SEQ ID NO: 739; 246bp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Corynebacterium glutamicum, and identifying a homologue of a gene derived
 CC from Corynebacterium glutamicum. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX

SQ Sequence 702 BP; 188 A; 240 C; 156 G; 118 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 702;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY      10 gtcgacgcgcagca 23
        |||||||
DB      102 gtcgacgcgcagca 115

RESULT  18
AAV20504/c
ID      AAV20504 standard; DNA: 719 BP.
XX
XX      AAV20504;
AC
XX      30-JUN-1998 (first entry)
DT
XX
XX      Conus geographus conantokin DNA.
DE
XX      Conantokin; predatory cone snail; treatment: neurologic disorder;
KW      psychiatric disorder; anticonvulsant; neuroprotective; analgesic;
KW      HIV infection; ophthalmic indication; memory; learning defect;
KW      cognitive defect; ss.
XX
OS      Conus geographus.
XX
XX      Key      Location/Qualifiers
FH      CDS      110..412
FT      /*tag= a
FT      /product= conantokin
XX
XX      W09803541-A1.
XX
XX      29-JAN-1998.
XX
XX      21-JUL-1997; 97WO-US12618.
XX
XX      22-JUL-1996; 96US-0684742.
XX
XX      (COGN-) COGNETIX INC.
XX      (UTAH-) UNIV UTAH RES FOUND.
XX
XX      Abogadie FC, Colledge C, Cruz LJ, Hillyard DR, Jimenez E;
PI      Layer RT, McCabe RT, Oliveira BM, Rivier JE, Shen GS;
XX      Walker C, Zhou L;
XX
XX      WPI: 1998-120694/11.
DR      P-PSDB: AAW48210.
XX
XX      New conantokin peptide(s) - useful for e.g. treating neurologic or
PT      psychiatric disorders, or the management of pain
PS
XX      Claim 19: Pages 79-80; 122pp; English.
XX
XX      The present sequence encodes Conus geographus conantokin, peptide
CC      derivatives of which can be used to treat neurologic and
CC      psychiatric disorders, e.g. as an anticonvulsant, neuroprotective
CC      or analgesic agent. Neurologic and psychiatric disorders include
CC      epilepsy, convulsions, neurotoxic injury (associated with
CC      conditions of hypoxia, anoxia or ischaemia, which typically follow
CC      stroke, cerebrovascular accident, brain or spinal cord trauma,
CC      myocardial infarct, physical trauma, drowning, suffocation,
CC      perinatal asphyxia or hypoglycaemic events), neurodegeneration
CC      (associated with Alzheimer's disease, senile dementia, Amyotrophic
CC      Lateral Sclerosis, Multiple Sclerosis, Parkinson's disease,
CC      Huntington's disease, Down's Syndrome, Korsakoff's disease,
CC      schizophrenia, AIDS dementia, multi-infarct dementia, Binswanger
CC      dementia and neuronal damage associated with uncontrolled
CC      seizures), chemical toxicity (such as addiction, and morphine,
CC      opiate, opioid and barbiturate tolerance), pain (acute, chronic,
CC      migraine), anxiety, major depression, manic-depressive illness,
CC      obsessive-compulsive disorder, schizophrenia and mood disorders
CC      (such as bipolar disorder, unipolar depression, dysthymia and
CC      seasonal affective disorder) and dystonia (movement disorder).
CC      sleep disorder, muscle relaxation and urinary incontinence. The
CC      peptide can also be used to treat HIV infection, ophthalmic
CC      indication and memory, learning or cognitive defects.

```

```

XX
XX      Sequence 719 BP; 217 A; 171 C; 149 G; 182 T; 0 other;
SQ
XX
XX      Query Match      56.0%; Score 14; DB 19; Length 719;
XX      Best Local Similarity 100.0%; Pred. No. 1.1e+02;
XX      Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      OY      7 caagtcgacgcgcga 20
XX      |||||||
DB      118 CAGGTGATCGCACA 105

RESULT  19
AAV17130/c
ID      AAV17130 standard; DNA: 719 BP.
XX
XX      AAV17130;
AC
XX      30-JUN-1998 (first entry)
DT
XX
XX      Conus geographus conantokin DNA.
DE
XX      Conantokin; predatory cone snail; treatment: neurologic disorder;
KW      psychiatric disorder; anticonvulsant; neuroprotective; analgesic;
KW      HIV infection; ophthalmic indication; memory; learning defect;
KW      cognitive defect; ss.
XX
OS      Conus geographus.
XX
XX      Key      Location/Qualifiers
FH      CDS      110..412
FT      /*tag= a
FT      /product= conantokin
XX
XX      W09803189-A1.
XX
XX      29-JAN-1998.
XX
XX      21-JUL-1997; 97WO-US12652.
XX
XX      06-DEC-1996; 96US-0762377.
XX      22-JUL-1996; 96US-0684750.
XX
XX      (COGN-) COGNETIX INC.
XX
XX      Layer RT, McCabe RT, Zhou L;
PI
XX
XX      WPI: 1998-120469/11.
DR      P-PSDB: AAW49989.
XX
XX      Example 4: Pages 79-80; 122pp; English.
XX
XX      The present sequence encodes Conus geographus conantokin, peptide
CC      derivatives of which can be used to treat neurologic and
CC      psychiatric disorders, e.g. as an anticonvulsant, neuroprotective
CC      or analgesic agent. Neurologic and psychiatric disorders include
CC      epilepsy, convulsions, neurotoxic injury (associated with
CC      conditions of hypoxia, anoxia or ischaemia, which typically follow
CC      stroke, cerebrovascular accident, brain or spinal cord trauma,
CC      myocardial infarct, physical trauma, drowning, suffocation,
CC      perinatal asphyxia or hypoglycaemic events), neurodegeneration
CC      (associated with Alzheimer's disease, senile dementia, Amyotrophic
CC      Lateral Sclerosis, Multiple Sclerosis, Parkinson's disease,
CC      Huntington's disease, Down's Syndrome, Korsakoff's disease,
CC      schizophrenia, AIDS dementia, multi-infarct dementia, Binswanger
CC      dementia and neuronal damage associated with uncontrolled
CC      seizures), chemical toxicity (such as addiction, and morphine,
CC      opiate, opioid and barbiturate tolerance), pain (acute, chronic,

```

CC migraine, anxiety, major depression, manic-depressive illness,
 CC obsessive compulsive disorder, schizophrenia and mood disorders
 CC (such as bipolar disorder, unipolar depression, dysthymia and
 CC seasonal affective disorder) and dystonia (movement disorder),
 CC sleep disorder, muscle relaxation and urinary incontinence. The
 CC peptide can also be used to treat HIV infection, ophthalmic
 CC indication and memory, learning or cognitive defects.
 CC
 SQ Sequence 719 BP; 217 A; 171 C; 149 G; 182 T; 0 other;
 Query Match 56.0%; Score 14; DB 19; Length 719;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 caggtcagtcgcca 20
 DB 118 CAGTGCATCGCCA 105
 RESULT 20
 AAH06472/c
 ID AAH06472 standard; cDNA; 844 BP.
 AC AAH06472;
 XX
 XX 26-JUN-2001 (first entry)
 DT
 XX Human cDNA clone (5'-primer) SEQ ID NO:3307.
 DE
 XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
 KW
 XX Homo sapiens.
 OS
 XX EP1074617-A2.
 PN
 XX 07-FEB-2001.
 PD
 XX 28-JUL-2000; 2000EP-0116126.
 PF
 XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 XX (HELI-) HELIX RES INST.
 PA
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 XX
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PT
 XX Claim 1; SEQ ID 3307; 2537pp + CD ROM; English.
 PS
 XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 CC
 SQ Sequence 844 BP; 195 A; 235 C; 247 G; 163 T; 4 other;
 Query Match 56.0%; Score 14; DB 22; Length 844;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ggcgagcaggtgca 14
 DB 399 GCGCAGCAGTGCA 386
 RESULT 21
 AAH48556
 ID AAH48556 standard; DNA; 943 BP.
 AC AAH48556;
 XX
 XX 20-SEP-2001 (first entry)
 DT
 XX Human fascin DNA fragment SEQ ID 6.
 DE
 XX
 XX Fascin; regulatory sequence; human; dendritic cell; antiviral; tumor;
 KW antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
 KW immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
 KW Creutzfeldt-Jakob disease; Alzheimer's disease; gene therapy;
 KW autoimmune disease; transplant rejection; ds.
 XX
 XX Homo sapiens.
 OS
 XX WO200151631-A2.
 PN
 XX 19-JUL-2001.
 PD
 XX 12-JAN-2001; 2001WO-BP00362.
 PF
 XX 13-JAN-2000; 2000DE-1001169.
 PR 02-MAR-2000; 2000DE-1010188.
 PR
 XX (RESK/) RESKE-KUNZ A.
 PA (ROSS/) ROSS X.
 PA (ROSS/) ROSS R.
 PA (BROS/) BROS M.
 XX
 XX Reske-Kunz A, Ross X, Bros R, Bros M;
 PI
 XX WPI: 2001-451858/48.
 DR
 XX New regulatory sequences from the fascin gene, useful for providing
 PT dendritic cell-specific expression of e.g. antigens, e.g. for
 PT vaccination against tumors and infections -
 PT
 XX Claim 1c; Page 87; 117pp; German.
 PS
 XX This invention describes novel regulatory sequences (A) derived from
 CC human fascin that provide specific expression in dendritic cells (DC) and
 CC which have antiviral, antibacterial, antifungal, antiparasitic,
 CC anti-allergic, neurological, immunomodulatory and apoptotic activity.
 CC (A) are used to regulate expression of antigens, immunoregulators,
 CC antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors
 CC and host cells that contain (A) are useful: (i) in vaccines against
 CC viruses, bacteria, fungi, parasites, tumors, allergens and plaques in
 CC Creutzfeldt-Jakob and Alzheimer's disease; and (ii) for gene therapy of

CC tumors, allergies, infections, autoimmune diseases and transplant
CC rejection. They can also be provide specific expression of antigens and
CC immunoregulators in DC; for isolation and identification of cell factors
CC and cis-elements from regulatory sequences that mediate DC-specific
CC expression; to determine the degree of maturity of DC and to block
CC transcription factors, by providing binding sites in DC. (A) provide
CC DC-specific expression of nucleic acid under their control, allowing a
CC more specific regulation of the immune response and eliminating the long
CC and laborious purification of DC (since a complete leucocyte population
CC may be transformed), including transformation in vitro. This sequence
CC represents a DNA fragment of human fascin which is described in the
CC invention.

CC Sequence 943 BP; 135 A; 307 C; 340 G; 161 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 943;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 gcacgcacgacatt 25
DB 111 gcatgcacgacatt 124

RESULT 22
AAH64755/C
ID AAH64755 standard; cDNA; 1116 BP.

XX AAH64755;

DT 11-SEP-2001 (first entry)

XX Human secreted protein cDNA, SEQ ID NO: 31.

DE Human; secreted protein; gene therapy; vaccine; treatment; diagnosis;

KW GENSET; ss.

CS Homo sapiens.

XX MO200142451-A2.

PD 14-JUN-2001.

PF 07-DEC-2000; 2000MO-IB01938.

XX 08-DEC-1999; 990S-0169629.

PR 06-MAR-2000; 2000US-0187470.

PA (GENSET) GENSET.

PI Dumas Milne Edwards J, Bouqueleret L, Jobert S;

DR WPI: 2001-367870/38.

DR P-PSDB; AAG89152.

PT Full length GENSET human nucleic acids encoding potentially secreted
PT proteins, useful in gene therapy and vaccination against a variety of
PT diseases, and for diagnosis of those diseases -

XX Claim 7; Page 590-591; 921pp; English.

XX The invention relates to full length GENSET human nucleic acids encoding
XX potentially secreted proteins. The nucleic acids and the polypeptides
XX they encode may be used in the prevention, treatment and diagnosis of
XX diseases associated with inappropriate GENSET gene expression. For
XX example, they be used to treat disorders associated with decreased
XX GENSET gene expression by rectifying mutations or deletions in a
XX patient's genome that affect the activity of GENSET or by supplementing
XX the patient's own production of GENSET polypeptides. Conversely,
XX antisense nucleic acid molecules may be administered to down regulate
XX GENSET expression by binding with the cells' own genes and preventing
XX their expression. The sense and antisense nucleic acids may also be

CC used as DNA probes in diagnostic assays to detect and quantitate the
CC presence of similar nucleic acid sequences in samples, and hence to
CC determine which patients may be in need of restorative therapy.
CC The GENSET polypeptides may be used as antigens in the production of
CC antibodies and in assays to identify modulators (agonists and
CC antagonists) of GENSET polypeptide expression and activity. The
CC present sequence is a GENSET nucleic acid of the invention.

CC Sequence 1116 BP; 324 A; 239 C; 283 G; 270 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1116;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcacgaggtgacac 16
DB 343 GCACGAGGTGACATC 330

RESULT 23
AAH66288/C
ID AAH66288 standard; DNA; 1131 BP.

XX AAH66288;

DT 26-SEP-2001 (first entry)

XX C glutamicum coding sequence fragment SEQ ID NO: 1323.

DE Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis; ds.

XX Corynebacterium glutamicum.

OS EP1108790-A2.

PN 20-JUN-2001.

PD 18-DEC-2000; 2000EP-0127688.

PF 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOGYO KK.

PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

DR WPI: 2001-376931/40.

DR P-PSDB; AAG91069.

PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene

XX Claim 8; SEQ ID NO: 1323; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein
XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX are useful for identifying the mutation point of a gene derived from a
XX mutant of coryneform bacterium, measuring expression amount and
XX analysing the expression profile or expression pattern of a gene derived
XX from Coryneform bacterium, and identifying a homologue of a gene derived
XX from coryneform bacterium. Coryneform bacteria are useful for producing
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX particularly L-lysine. The present sequence is a nucleic acid described
XX in the exemplification of the invention.

XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX European Patent Office.

XX Sequence 1131 BP; 253 A; 359 C; 302 G; 217 T; 0 other;
SQ

Query Match 56.0%; Score 14; DB 22; Length 1131;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcagcaggtgca 14
|||||
DB 832 GCGACAGCGTGA 819

RESULT 24
AAC98831
ID AAC98831 standard; cDNA; 1190 BP.

XX AAC98831;
XX
XX 09-MAR-2001 (first entry)

DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:59.

XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
XX detection; diagnosis; identification; cytostatic; neuroprotective;
XX neurotropic; immunomodulatory; relaxant; contraceptive; gynaecological;
XX antinflammatory; cardiant; gene therapy; chromosome mapping;
XX linkage analysis; tissue identification; tissue typing; forensic;
XX neutral; immune system; muscular; reproductive; gastrointestinal;
XX pulmonary; cardiovascular; renal; proliferative; ss.

OS Homo sapiens.

PN WO200055320-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-0505989.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI: 2000-579444/54.

XX P-PSDB; AAB54066.

XX New nucleic acid that is a pancreatic cancer antigen for preventing,
XX treating, or ameliorating a medical condition, particular pancreatic
XX cancer, or for use in assays for diagnosing a pathological condition -
PS Claim 1; Page 537; 1379pp; English.

CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, neurotropic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiant and antinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.

CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,

CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.
XX

SQ Sequence 1190 BP; 269 A; 323 C; 329 G; 267 T; 2 other;

Query Match 56.0%; Score 14; DB 21; Length 1190;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 ggtgcatcgccagc 22
|||||
DB 1141 ggtgcatcgccagc 1154

RESULT 25
AAH48555
ID AAH48555 standard; DNA; 1240 BP.

XX AAH48555;
XX
XX 20-SEP-2001 (first entry)

DE Human fascin DNA fragment SEQ ID 5.

XX Fascin; regulatory sequence; human; dendritic cell; antiviral; tumor;
XX antifungal; antifungal; antiparasitic; anti-allergic; neurological;
XX immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
XX Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
XX autoimmune disease; transplant rejection; ds.

OS Homo sapiens.

PN WO200151631-A2.

XX 19-JUL-2001.

XX 12-JAN-2001; 2001WO-EP00362.

XX 13-JAN-2000; 2000DE-1001169.

XX 02-MAR-2000; 2000DE-1001188.

XX (RESK/) RESKE-KUNZ A.

XX (ROSS/) ROSS X.

XX (ROSS/) ROSS R.

XX (BROS/) BROS M.

XX Reske-Kunz A, Ross X, Ross R, Bros M;

XX WPI: 2001-451858/48.

XX New regulatory sequences from the fascin gene, useful for providing

XX dendritic cell-specific expression of e.g. antigens, e.g. for
XX vaccination against tumors and infections -
PS Claim 1c; Page 87; 117pp; German.

CC This invention describes novel regulatory sequences (A) derived from
CC human fascin that provide specific expression in dendritic cells (DC) and
CC which have antiviral, antibacterial, antifungal, antiparasitic,
CC anti-allergic, neurological, immunomodulatory and apoptotic activity.
CC (A) are used to regulate expression of antigens, immunoregulators,
CC antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors
CC and host cells that contain (A) are useful: (i) in vaccines against
CC viruses, bacteria, fungi, parasites, tumors, allergens and plaques in
CC Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of
CC tumors, allergies, infections, autoimmune diseases and transplant
CC rejection. They can also be provide specific expression of antigens and
CC immunoregulators in DC; for isolation and identification of cell factors
CC and cis-elements from regulatory sequences that mediate DC-specific
CC expression; to determine the degree of maturity of DC and to block
CC transcription factors, by providing binding sites in DC. (A) provide

CC DC-specific expression of nucleic acid under their control, allowing a
 CC more specific regulation of the immune response and eliminating the long
 CC and laborious purification of DC (since a complete leucocyte population
 CC may be transformed), including transformation in vitro. This sequence
 CC represents a DNA fragment of human fascin which is described in the
 CC invention.

XX Sequence 1240 BP: 191 A; 364 C; 466 G; 219 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1240;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 gcatgcacgacatt 25
 Db 408 gcatgcacgacatt 421

RESULT 26

AAH66089/C
 ID AAH66089 standard; DNA: 1278 BP.

AC AAH66089;

DT 26-SEP-2001 (first entry)

DE C glutamicum coding sequence fragment SEQ ID NO: 1124.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

XX organic acid synthesis; ds.

OS Corynebacterium glutamicum.

PN EP1108790-A2.

PD 20-JUN-2001.

PF 18-DEC-2000; 2000EP-0127688.

PR 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOWA) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI: 2001-376931/40.

DR P-PSDB: AAG90870.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene

XX Claim 8; SEQ ID NO: 1124; 246pp + Sequence Listing: English.

XX The present invention provides a number of nucleotide and protein

CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a

CC mutant of coryneform bacterium, measuring expression amount and

CC analysing the expression profile or expression pattern of a gene derived

CC from Coryneform bacterium, and identifying a homologue of a gene derived

CC from coryneform bacterium. Coryneform bacteria are useful for producing

CC amino acids, nucleic acids, vitamins, saccharides and organic acids,

CC particularly L-lysine. The present sequence is a nucleic acid described

CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the

CC European Patent Office.

SQ Sequence 1278 BP: 295 A; 355 C; 348 G; 280 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1278;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggcgcagcagtgca 14

Db 1077 GCGCAGCAGGTGCA 1064

RESULT 27

AAH68314
 ID AAH68314 standard; DNA: 1326 BP.

AC AAH68314;

DT 26-SEP-2001 (first entry)

DE C glutamicum coding sequence fragment SEQ ID NO: 3349.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

XX organic acid synthesis; ds.

OS Corynebacterium glutamicum.

PN EP1108790-A2.

PD 20-JUN-2001.

PF 18-DEC-2000; 2000EP-0127688.

PR 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOWA) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI: 2001-376931/40.

DR P-PSDB: AAG93095.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene

XX Claim 8; SEQ ID NO: 3349; 246pp + Sequence Listing: English.

XX The present invention provides a number of nucleotide and protein

CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a

CC mutant of coryneform bacterium, measuring expression amount and

CC analysing the expression profile or expression pattern of a gene derived

CC from Coryneform bacterium, and identifying a homologue of a gene derived

CC from coryneform bacterium. Coryneform bacteria are useful for producing

CC amino acids, nucleic acids, vitamins, saccharides and organic acids,

CC particularly L-lysine. The present sequence is a nucleic acid described

CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the

CC European Patent Office.

Query Match 56.0%; Score 14; DB 22; Length 1326;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcgcagcaggtgca 14
 |||
 Db 68 gcgcagcaggtgca 81

RESULT 28

AA256381/c
 ID AA256381 standard; DNA; 1344 BP.

XX AA256381;

DT 17-MAR-2000 (first entry)

DE Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:51.

XX Flagellin; fliC; antigen; detection; ds.

OS Escherichia coli.

PN W09961458-A1.

PD 02-DEC-1999.

PF 21-MAY-1999; 99WO-AU00385.

PR 21-MAY-1998; 98AU-0003634.

XX (UNSY) UNITV SYDNEY.

PI Reeves PR, Wang L;

XX WPI: 2000-072598/06.

PT Novel nucleic acid molecule useful for the detection of flagellated
 bacterial strains in food, faeces, etc.

PS Claim 3; Page 225; 245pp: English.

XX AA256331 to AA256398 represent nucleic acid molecules (1) encoding all
 or part of an Escherichia coli flagellin protein except a protein
 expressed by E. coli H1, H7, H12 or H48 type strains. The present
 invention also describes a method of detecting the presence of E. coli
 of a particular H serotype in a sample, comprising specifically
 hybridising a nucleic acid, preferably at least a pair, derived from a
 flagellating gene, specific for a particular flagellin gene associated
 with the H serotype, to any E. coli in the sample which contain the gene,
 and detecting any hybridised molecules, identifying the presence of that
 serotype in the sample. (1) are useful for: (1) detecting the presence
 of E. coli of H serotype in a sample by hybridising at least one or a
 pair of (1) to any E. coli in the sample and detecting the hybridised
 nucleic acid molecules; and (2) for detecting the presence of both O
 and H-serotypes of E. coli by hybridising at least one or a pair of (1)
 to any E. coli present in the sample and detecting the hybridised
 nucleic acid molecules. (1) is particularly useful for detecting the
 combination of O and H antigen. Hybridised (1) when using at least one
 (1) is detected by southern blot analysis and, when using a pair of (1),
 is detected by polymerase chain reaction (PCR). AA256399 to AA256420
 represent primers used in the exemplification of the present invention.

XX Sequence 1344 BP; 362 A; 337 C; 322 G; 323 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1344;

Best Local Similarity 100.0%; Pred. NO. 1.1e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 gcacgcagcagcatt 25
 |||
 Db 590 GCATGCCAGCAGCATT 577

RESULT 29
 AA158241/c

ID AA158241 standard; cDNA; 1431 BP.

XX AA158241;

DT 22-OCT-2001 (first entry)

DE Human polynucleotide SEQ ID NO 444.

XX Human; nootropic; immunosuppressant; cytosolic; gene therapy; cancer;

KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW chemokine; thrombolytic; drug screening; arthritis; inflammation;

XX leukaemia; ss.

XX Homo sapiens.

PN W020015312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

DR P-PSDB: AAM39085.

PT Novel nucleic acids and polypeptides, useful for treating disorders

XX such as central nervous system injuries -

PS Claim 1; SEQ ID NO 444; 10078bp: English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and

XX the encoded polypeptides (AAM38642-AAM44213) with nootropic.

XX immunosuppressant and cytosolic activity. The polynucleotides are useful

XX in gene therapy. A composition containing a polypeptide or polynucleotide

XX of the invention may be used to treat diseases of the peripheral nervous

XX system, such as peripheral nervous injuries, peripheral neuropathy and

XX localised neuropathies and central nervous system diseases, such as

XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

XX utilisation of the activities such as: immune system suppression,

XX activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,

XX assays for receptor activity, arthritis and inflammation, leukaemias and

XX C.N.S disorders.

XX Note: The sequence data for this patent did not form part of the printed

XX specification.

SQ Sequence 1431 BP; 414 A; 277 C; 309 G; 431 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1431;

Best Local Similarity 100.0%; Pred. NO. 1.1e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcgcagcagtcac 16
 |||
 Db 493 GCAGCAGCAGTCATC 480

XX Sequence 1934 BP: 344 A: 559 C: 647 G: 384 T: 0 other:
SQ

Query Match 56.0%; Score 14; DB 22; Length 1934;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 gcatcgccagcat 25
|
DB 1102 gcatcgccagcat 1115

RESULT 34
AAA69762/c
ID AAA69762 standard; cDNA: 2017 BP.

XX AAA69762;

XX 07-NOV-2000 (first entry)

XX Human ovarian carcinoma antigen polynucleotide SEQ ID NO:72.

XX Human: ovarian carcinoma; ovarian cancer; therapy; diagnosis;
XX tumour antigen; identification; cytostatic; gene therapy; vaccine; ss.

XX Homo sapiens.

XX WO200036107-A2.

XX 22-JUN-2000.

XX 17-DEC-1999; 99WO-US30270.

XX 17-DEC-1998; 98US-0215681.

XX 17-DEC-1998; 98US-0216003.

XX 23-JUN-1999; 99US-0338933.

XX 24-SEP-1999; 99US-0404879.

XX (CORI-) CORIAX CORP.

XX Mitcham JL, King GE, Algate PA, Fridakis TN;

XX WPI: 2000-431589/37.

XX Immunogenic portion of an ovarian carcinoma protein and the nucleic
XX acid encoding it, useful for the diagnosis, prevention and treatment of
XX cancer, preferably ovarian cancer.

XX Claim 1; Fig 2: 299pp; English.

XX The present invention describes an isolated polypeptide comprising an
XX immunogenic portion of an ovarian carcinoma protein (or its variants).

XX Ovarian carcinoma proteins, and polynucleotides encoding them, have
XX cytostatic activity and can be used in gene therapy and vaccines.

XX Ovarian carcinoma polypeptides, nucleic acids, antibodies and vaccines
XX are useful for the prevention, diagnosis and treatment of cancer.

XX preferably ovarian cancer. AAA69691 to AAA70077 and AAB12552
XX represent human ovarian carcinoma polynucleotides and proteins used in
XX the exemplification of the present invention.

XX Sequence 2017 BP: 673 A: 440 C: 462 G: 442 T: 0 other;

Query Match 56.0%; Score 14; DB 21; Length 2017;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 agtgcacgcacg 21
|
DB 964 AGTGCATCGCCAC 951

RESULT 35
AAT30684/c
ID AAT30684 standard; DNA: 2193 BP.

XX AAT30684;

XX 11-SEP-1996 (first entry)

XX Kaposi's sarcoma associated herpesvirus ORF22.

XX Kaposi's sarcoma associated herpesvirus; KSHV; gamma-2 herpesvirus;
XX therapy; diagnosis; antisense; triplex; vaccine; AIDS; ss.

XX Kaposi's sarcoma associated herpesvirus.

XX WO9615779-A1.

XX 30-MAY-1996.

XX 21-NOV-1995; 95WO-US15138.

XX 11-APR-1995; 95US-0420235.

XX 21-NOV-1994; 94US-0343101.

XX (UYCO) UNIV COLUMBIA NEW YORK.

XX Chang Y, Moore PS;

XX WPI: 1996-268320/27.

XX P-PSDB; AAR97832.

XX Herpes virus associated with Kaposi's sarcoma - also definitive DNA
XX sequences, useful for diagnosis of and to develop prods. for
XX treatment of Kaposi's sarcoma

XX Claim 17: Page 198-201; 277pp; English.

XX Lambda clone KS5 (AAT30681) has 17 protein-coding regions (AAT30682-98),
XX 15 of which are complete ORFs, including ORF22 (AAT30684) (nts 2093-
XX 4285), which codes for glycoprotein-H (AAR97832). KS5 is a fragment
XX of a newly identified human gamma-2 herpesvirus associated with

XX Kaposi's sarcoma (KS). The ORFs were identified by sequence and
XX positional homology to known herpesvirus sequences. Probes that bind
XX specifically to the KS associated herpesvirus can be used for KS

XX diagnosis. Antisense or triplex oligonucleotides are useful for
XX prophylaxis or treatment of KS, and the protein products (see also
XX AAR97830-46) of the 17 open reading frames are useful as vaccines.

XX Sequence 2193 BP: 510 A: 597 C: 536 G: 550 T: 0 other;

Query Match 56.0%; Score 14; DB 17; Length 2193;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgcacgcat 24
|
DB 46 TGCATCGCCAGCAT 33

RESULT 36
AAT16814/c
ID AAT16814 standard; DNA: 2193 BP.

XX AAT16814;

XX 13-AUG-1996 (first entry)

XX Kaposi's sarcoma associated herpesvirus ORF22.

XX Kaposi's sarcoma; KSHV; lymphoma; AIDS; vaccine; diagnosis; therapy;
XX glycoprotein H; ss.

XX

XX WPI: 2001-376931/40.
DR P-PSDB: AAC90181.
XX
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
PS
PS Claim 8; SEQ ID NO: 435; 246bp + Sequence Listing: English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from corynebacterium bacterium, and identifying a homologue of a gene derived
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX Sequence 2265 BP; 513 A; 769 C; 556 G; 427 T; 0 other;
SQ
XX
XX Query Match 56.0%; Score 14; DB 22; Length 2265;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 3 gcagcaggtgcatc 16
|||||
DB 637 GCAGCAGGTGCATC 624
XX
XX RESULT 39
AAAF61014/C
ID AAF61014 standard; DNA; 2397 BP.
XX
XX AAF61014;
XX
XX 16-MAY-2001 (first entry)
XX
XX P. putida KT2440-associated DNA ORF04125.
XX
XX Transgenic plant; detection; probe; amplification; vaccine carrier;
KW microbial production strain; biological remediation; ds.
XX
XX Pseudomonas putida.
OS
XX DEL19935088-A1.
XX
XX 01-FEB-2001.
XX
XX 27-JUL-1999; 99DE-1035088.
XX
XX 27-JUL-1999; 99DE-1035088.
XX
XX (TIGR-) TIGR INST GENOMIC RES.
PA (GIB) GES BIOLOGISCHES FORSCHUNG MBH.
PA (GIB) GES BIOLOGISCHES FORSCHUNG MBH.
PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
PA (MED1-) MEDIZINISCHE HOCHSCHULE HANNOVER.
XX
XX WPI: 2001-192469/20.
XX
XX New DNA sequences specific for Pseudomonas putida KT2440, useful as
PT safe genetic engineering host, allow detection in presence of other
PT related bacteria -
XX
XX Claim 1a; Page 62-64; 158pp; German.

XX
XX This invention describes novel DNA sequences (1) for specific detection
CC of Pseudomonas putida KT2440. The invention also describes (1)
CC recombinant expression vector containing (1); (2) prokaryotic or
CC eukaryotic cells transformed or transfected with (1) or the vector of
CC (1); (3) production of expression products by culturing cells of (2);
CC (4) expression products, or their fragments, of (1) and synthetic
CC proteins or peptides with the same sequences (A); (5) poly- or
CC mono-clonal antibodies (Ab) that react specifically with (A); (6)
CC hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic
CC plants that contain transformed or transfected cells of (2); (8)
CC detecting KT2440 using a labeled (1) or Ab as probe, and (9) DNA chips
CC carrying one or more (1). (1), and their fragments, are used as probes
CC to detect and isolate full-length cDNAs and/or to amplify such cDNAs by
CC polymerase chain reaction, and for production of transgenic plants. (1),
CC or antibodies that recognize their expression products, are used for
CC detecting the presence of KT2440, particularly in presence of other,
CC even closely related, bacteria. KT2440 is one of the bacteria classified
CC as safe, by the National Institutes of Health, for genetic engineering
CC work, e.g. as microbial production strains, for biological remediation
CC and as vaccine carriers. (1) are exclusive to KT2440 with no significant
CC homology with sequences in other bacteria (specifically the closely
CC related pathogen P. aeruginosa). Compared with other 'safe' bacteria, it
CC has greater catabolic activity and better survival in, and adaptation to,
CC the rhizosphere and soil.
XX
XX Sequence 2397 BP; 488 A; 792 C; 708 G; 406 T; 3 other;
SQ
XX
XX Query Match 56.0%; Score 14; DB 22; Length 2397;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 gcgcagcaggtgca 14
|||||
DB 723 GCAGCAGGTGCATC 710
XX
XX RESULT 40
AAH14473/C
ID AAH14473 standard; cDNA; 2497 BP.
XX
XX AAH14473;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human cDNA sequence SEQ ID NO:11969.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
KW Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
XX 27-AUG-1999; 99JP-0300253.
XX
XX 11-JAN-2000; 2000JP-0118776.
XX
XX 02-MAY-2000; 2000JP-0183767.
XX
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI: 2001-318749/34.
XX
XX primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8: SEQ ID 11969; 2537bp + CD ROM; English.
 CC The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide which comprises a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC in gene therapy. The primer sets can be used in antisense therapy and
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 2497 BP; 538 A; 711 C; 702 G; 546 T; 0 other;
 Query Match
 Best Local Similarity 56.0%; Score 14; DB 22; Length 2497;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ggcgcagcagcgca 14
 ||||||||||||
 Db 399 GCGCAGCAGCGTCA 366
 RESULT 41
 AAH48551
 ID AAH48551 standard; DNA; 3069 BP.
 XX
 AC AAH48551;
 XX
 DT 20-SEP-2001 (first entry)
 XX
 DE Human fascin DNA fragment SEQ ID 1.
 XX
 DE Fascin; regulatory sequence; human; dendritic cell; antiviral; tumor;
 KW antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
 KW immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
 KW Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
 KW autoimmune disease; transplant rejection; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200151631-A2.
 XX
 PD 19-JUL-2001.
 XX
 PF 12-JAN-2001; 2001WO-EP00362.
 XX
 PR 13-JAN-2000; 2000DE-1001169.
 PR 02-MAR-2000; 2000DE-1010188.
 XX
 PA (RESK/) RESKE-KUNZ A.
 PA (ROSS/) ROSS X.
 PA (ROSS/) ROSS R.
 PA (BROS/) BROS M.
 XX
 PI Reske-Kunz A, Ross X, Ross R, Bros M;

XX
 DR WPI; 2001-451858/48.
 XX
 PT New regulatory sequences from the fascin gene, useful for providing
 PT dendritic cell-specific expression of e.g. antigens, e.g. for
 PT vaccination against tumors and infections -
 XX
 PS Claim 1a: Page 84-85; 117pp; German.
 CC
 CC This invention describes novel regulatory sequences (A) derived from
 CC human fascin that provide specific expression in dendritic cells (DC) and
 CC which have antiviral, antibacterial, antifungal, antiparasitic,
 CC anti-allergic, neurological, immunomodulatory and apoptotic activity.
 CC (A) are used to regulate expression of antigens and apoptotic activity.
 CC antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors
 CC and host cells that contain (A) are useful: (i) in vaccines against
 CC viruses, bacteria, fungi, parasites, tumors, allergens and plaques in
 CC Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of
 CC tumors, allergies, infections, autoimmune diseases and transplant
 CC rejection. They can also be provide specific expression of cell factors
 CC and immunoregulators in DC; for isolation and identification of antigens and
 CC and cis-elements from regulatory sequences that mediate DC-specific
 CC expression; to determine the degree of maturity of DC and to block
 CC transcription factors, by providing binding sites in DC. (A) provide
 CC DC-specific expression of nucleic acid under their control, allowing a
 CC more specific regulation of the immune response and eliminating the long
 CC and laborious purification of the immune response and eliminating the long
 CC may be transformed), including transformation in vitro. This sequence
 CC represents a DNA fragment of human fascin which is described in the
 CC invention.
 XX
 SQ Sequence 3069 BP; 607 A; 867 C; 993 G; 602 T; 0 other;
 Query Match
 Best Local Similarity 56.0%; Score 14; DB 22; Length 3069;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 12 gcatgcagcagcatt 25
 ||||||||||||
 Db 2237 gcatgcagcagcatt 2250
 RESULT 42
 AA239011/C
 ID AA239011 standard; cDNA; 3593 BP.
 XX
 AC AA239011;
 XX
 DT 28-FEB-2000 (first entry)
 XX
 DE Mouse Eae2 coding sequence.
 XX
 DE Mouse; murine; Eae2; endocytosis; vesicular trafficking;
 KW regulation; actin cytoskeleton; detection; cancer; infection;
 KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
 KW antiproliferative; antiviral; ss.
 XX
 OS Mus sp.
 XX
 PN WO9955728-A2.
 XX
 PD 04-NOV-1999.
 XX
 PF 27-APR-1999; 99WO-CA00375.
 XX
 PR 27-APR-1998; 98CA-2230201.
 PR 05-FEB-1999; 99US-0118739.
 XX
 PA (HSCR-) HSC RES & DEV LP.
 XX
 PI Egan SE, Wang W, Sengar A;

DR WPI: 2000-052802/04.
 P-PSDB: AAY57445.
 XX
 PT New nucleic acid encoding Esei and 2 proteins, involved in regulation
 of endocytosis, used e.g. for treating cancer or preventing viral
 PT infection -
 PS Claim 25; Page 46-48; 99pp; English.
 XX
 CC The present sequence encodes mouse Esei2. The present invention
 CC specifically describes mammalian Esei and 2 proteins (I) and their splice
 CC variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I)
 CC are involved in regulation of clathrin-mediated endocytosis (as a complex
 CC with Esp15 protein), vesicular trafficking and actin cytoskeleton.
 CC Generally (I) (or its (ant)agonists, mimetics, fragments and inactive
 CC mutants): (i)-specific antibodies (Ab); sequences antisense to the (I)
 CC polynucleotide; agents that downregulate expression of Ese genes or
 CC antagonists of an Ese binding partner are used to treat diseases
 CC associated with undesirable endocytosis and resulting changes in cellular
 CC function. Particularly overexpression of Esei is used to block
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while
 CC administration of (I) or Ab are used to suppress abnormal proliferation of
 CC cells that can be stimulated to proliferate by a growth factor receptor;
 CC and similar compounds (also inactive Ese mutants) can be used to prevent
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell
 CC cultures, by forming an Ese-Esp15 complex, then binding dynam to the
 CC complex. Generally conditions that can be treated include cancer;
 CC abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission.
 CC
 SQ Sequence 3593 BP; 1171 A; 783 C; 850 G; 789 T; 0 other:
 Query Match 56.0%; Score 14; DB 21; Length 3593;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 8 aggtgcatgcgcagc 21
 |||||
 DB 923 AGTGCATGCAGCAG 910
 RESULT 43
 ID AAH48581 standard; DNA; 4100 BP.
 AC AAH48581;
 XX
 DT 20-SEP-2001 (first entry)
 XX
 DE Human fascin DNA fragment SEQ ID 33.
 XX
 KW Fascin: regulatory sequence; human; dendritic cell; antiviral; tumor;
 KW antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
 KW immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
 KW Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
 KW autoimmune disease; transplant rejection; ds.
 XX
 OS Homo sapiens.
 XX
 PN MO200151631-A2.
 PD 19-JUL-2001.
 XX
 PF 12-JAN-2001; 2001WO-EP00362.
 XX
 PR 13-JAN-2000; 2000DE-1001169.
 PR 02-MAR-2000; 2000DE-1010188.
 XX
 PA (RESK/) RESKE-KUNZ A.
 PA (ROSS/) ROSS X.
 PA (ROSS/) ROSS R.

PA (BROS/) BROS M.
 XX
 PI Reske-Kunz A, Ross X, Ross R, Bros M;
 XX
 DR WPI: 2001-451858/48.
 XX
 PT New regulatory sequences from the fascin gene, useful for providing
 PT dendritic cell-specific expression of e.g. antigens, e.g. for
 PT vaccination against tumors and infections -
 PS Example 1; Page 102-103; 117pp; German.
 XX
 CC This invention describes novel regulatory sequences (A) derived from
 CC human fascin that provide specific expression in dendritic cells (DC) and
 CC which have antiviral, antibacterial, antifungal, antiparasitic,
 CC anti-allergic, neurological, immunomodulatory and apoptotic activity.
 CC (A) are used to regulate expression of antigens, immunoregulators,
 CC antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors
 CC and host cells that contain (A) are useful: (i) in vaccines against
 CC Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of
 CC tumors, allergies, infections, autoimmune diseases and transplant
 CC rejection. They can also be provide specific expression of cell factors
 CC and immunoregulators in DC; for isolation and identification of cell factors
 CC and cis-elements from regulatory sequences that mediate DC-specific
 CC expression; to determine the degree of maturity of DC and to block
 CC transcription factors, by providing binding sites in DC. (A) provide
 CC DC-specific expression of nucleic acid under their control, allowing a
 CC more specific regulation of the immune response and eliminating the long
 CC and laborious purification of DC (since a complete leucocyte population
 CC may be transformed), including transformation in vitro. This sequence
 CC represents a DNA fragment of human fascin which is described in the
 CC invention.
 SQ Sequence 4100 BP; 769 A; 1251 C; 1339 G; 741 T; 0 other:
 Query Match 56.0%; Score 14; DB 22; Length 4100;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 12 gcatgcgcagcatt 25
 |||||
 DB 2237 gcatgcgcagcatt 2250
 RESULT 44
 ID AA081778 standard; DNA; 4182 BP.
 AC AA081778;
 XX
 DT 09-AUG-1995 (first entry)
 XX
 DE DNA encoding cytohesin protein.
 XX
 KW Cytohesin protein; diagnostic probe; unusual codon usage;
 KW immunosassay; antigen; ds.
 XX
 OS Mycoplasma gallisepticum.
 XX
 FH Key Location/Qualifiers
 FT CDS 713..4081
 FT /tag= a
 FT /product= cytohesin protein
 FT misc_difference 887..889
 FT /*tag= b
 FT /codon= seq:iga, aa:trp
 FT misc_difference 998..1000
 FT /*tag= c
 FT /codon= seq:iga, aa:trp
 FT misc_difference 1484..1486
 FT /*tag= d

FT	misc_difference	/codon=seq:cta, aa:llys
FT		/*tag= e
FT		/codon=seq:tta, aa:llys
FT	misc_difference	1559..1561
FT		/*tag= f
FT		/codon=seq:tta, aa:llys
FT	misc_difference	1592..1594
FT		/*tag= g
FT		/codon=seq:tta, aa:llys
FT	misc_difference	1676..1678
FT		/*tag= h
FT		/codon=seq:tga, aa:trp
FT	misc_difference	1715..1717
FT		/*tag= i
FT		/codon=seq:cta, aa:llys
FT	misc_difference	1853..1855
FT		/*tag= j
FT		/codon=seq:tga, aa:trp
FT	misc_difference	1890..1892
FT		/*tag= k
FT		/codon=seq:tga, aa:trp
FT	misc_difference	1904..1906
FT		/*tag= l
FT		/codon=seq:tta, aa:llys
FT	misc_difference	1961..1963
FT		/*tag= m
FT		/codon=seq:tga, aa:trp
FT	misc_difference	2132..2134
FT		/*tag= n
FT		/codon=seq:cta, aa:llys
FT	misc_difference	2237..2239
FT		/*tag= o
FT		/codon=seq:tga, aa:trp
FT	misc_difference	2273..2275
FT		/*tag= p
FT		/codon=seq:tga, aa:trp
FT	misc_difference	2417..2419
FT		/*tag= q
FT		/codon=seq:tga, aa:trp
FT	misc_difference	2504..2506
FT		/*tag= r
FT		/codon=seq:tga, aa:llys
FT	misc_difference	2609..2611
FT		/*tag= s
FT		/codon=seq:tga, aa:trp
FT	misc_difference	2639..2641
FT		/*tag= t
FT		/codon=seq:tga, aa:trp
FT	misc_difference	2858..2860
FT		/*tag= u
FT		/codon=seq:tga, aa:trp
FT	misc_difference	3020..3022
FT		/*tag= v
FT		/codon=seq:tga, aa:trp
FT	misc_difference	3083..3085
FT		/*tag= w
FT		/codon=seq:tga, aa:trp
FT	misc_difference	3138..3140
FT		/*tag= x
FT		/codon=seq:tga, aa:trp
PN	US5378820-A.	
XX	03-JAN-1995.	
XX	09-NOV-1992.	92US-0973257.
XX	09-NOV-1992.	92US-0973257.
XX	09-NOV-1992.	92US-0973257.
PA	(DOHM.) DOHMS J E.	
PA	(KEEL.) KEELER C L.	

PI	Dohms JF, Keeler CL:
XX	
DR	WPI: 1995-051314/07.
DK	P-PSDB; AAR64927.
XX	
PT	Nucleic acid encoding cytoadhesin protein - used as a probe to
PR	diagnose Mycoplasma gallisepticum infection in poultry
XX	
PS	Claim 1; Column 7-16; 13pp. English.
XX	
CC	The sequence encodes a cytoadhesin protein from Mycoplasma
CC	gallisepticum. All or part of the sequence may be used as a probe
CC	for diagnosis of M. gallisepticum infection in poultry, e.g. turkey
CC	or fowl. The sequence has several UGA codons, which may be
CC	eliminated to avoid problems of expression in vitro due to irregular
CC	codon usage. The sequence may be cloned in a vector for recombinant
CC	cytoadhesin production, for use as an antigen in immunoassay formats
XX	
XX	
Sequence	4182 BP; 1456 A; 763 C; 746 G; 1217 T; 0 other:

```

Query Match      56.0%; Score 14; DB 16; Length 4182;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY	3	gcagcaggtgcac	16
Db	1260	GCAGCAGCTGCATC	1247

RESULT 45
AAZ39010/c
ID AAZ39010 standard; cDNA; 4625 BP.
xy

DT 28-FEB-2000 (first entry)
XY

Mouse Ees2 full length cDNA sequence.

NM mouse; murine; Ees1; Ees2; endocytosis; vesicular trafficking
 KW regulation; actin cytoskeleton; detection; cancer; infection;
 KW EH-domain and SH3-domain regulator of endocytosis; antineoplastic; antiviral; ss.
 KW antiproliferative; antiviral; ss.

Mus sp.

PN W09955728-A2.

PD 04-NOV-1999

PF 27-APR-1999; 99WO-CA00375.

PR 27-APR-1998; 98CA-2230201.

XX
XX

DT	XXXX	XXXX	XXXX
XX	XXXX	XXXX	XXXX

XX WPT : 3000-0E3803 404

DR P-PSDB; AAY57445.
XX

PT New nucleic acid end
PT of endocytosis

PJ Infection -
XX

PS Claim 25; Page 43-46
XX

The present sequence specifically describes

variants (Ese = EH-c

are involved in regulation of clathrin-mediated endocytosis (as a complex

CC with Esp15 protein), vesicular trafficking and actin cytoskeleton.
CC Generally (I) (or its (ant)agonists, mimetics, fragments and inactive
CC mutants); (I)-specific antibodies (Ab); sequences antisense to the (I)
CC polynucleotide; agents that downregulate expression of Ese genes or
CC antagonists of an Ese binding partner are used to treat diseases
CC associated with undesirable endocytosis and resulting changes in cellular
CC function. Particularly overexpression of Ese1 is used to block
CC clathrin-mediated endocytosis in vivo or in cell cultures, while
CC administration of (I) is used to promote endocytosis of selected cells.
CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
CC cells that can be stimulated to proliferate by a growth factor receptor;
CC and similar compounds (also inactive Ese mutants) can be used to prevent
CC viral infection. Endocytosis may also be regulated, in vivo or in cell
CC cultures, by forming an Ese-Esp15 complex, then binding dynamitin to the
CC complex. Generally conditions that can be treated include cancer;
CC abnormal cell division or migration; viral infection; or abnormal
CC receptor signalling, tissue development or synaptic transmission.
CC
XX
SQ Sequence 4625 BP; 1453 A; 1009 C; 1104 G; 1059 T; 0 other;

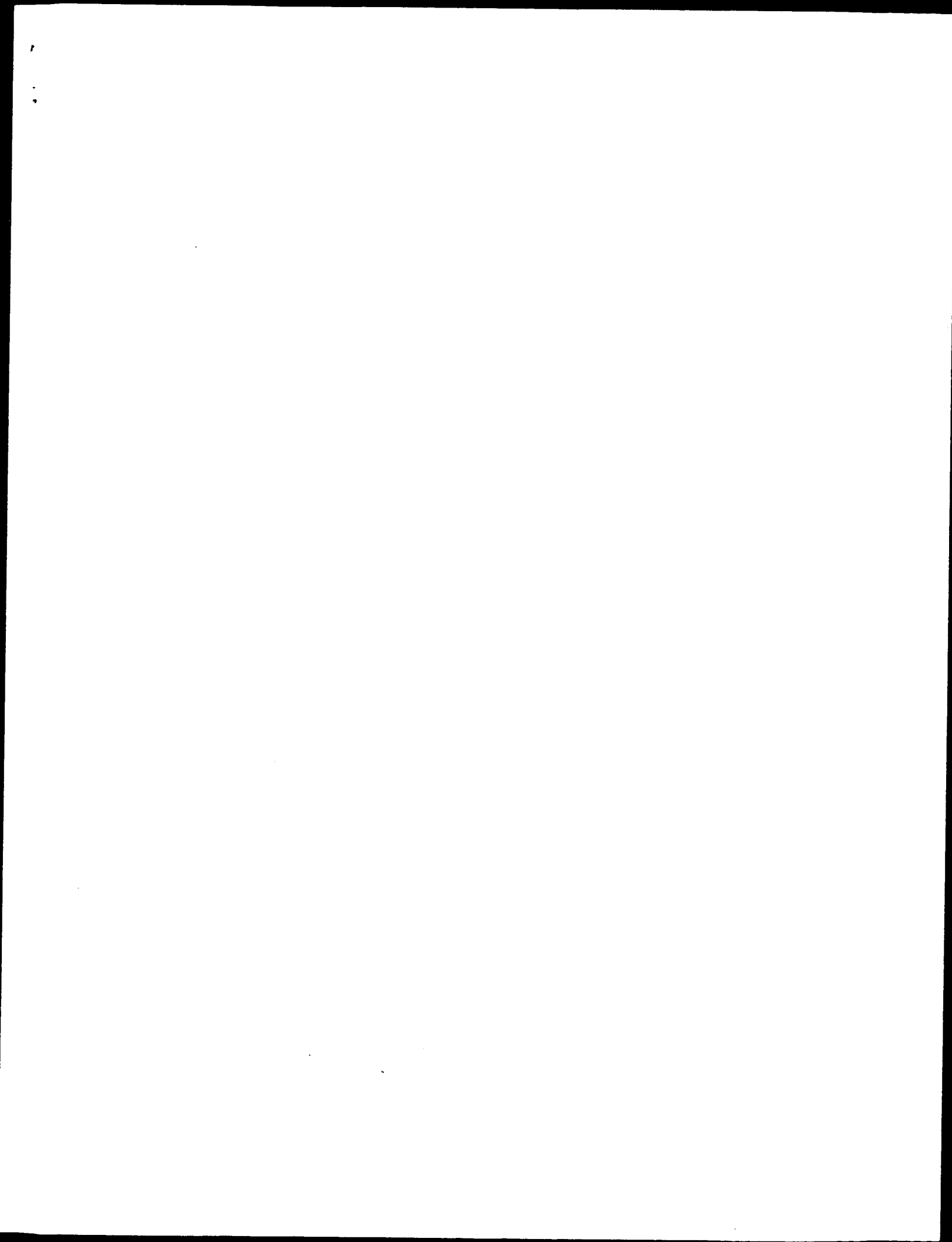
Query Match

56.0%; Score 14; DB 21; Length 4625;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 8 aggtgcatcgccag 21
|||||

DB 1173 AGGTGATCGCCAG 1160

Search completed: December 26, 2001, 14:16:43
Job time: 6995 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:52:37 : Search time 143.1 Seconds
(Without alignments)
39.566 Million cell updates/sec

Title: US-09-396-196f-5

Perfect score: 25

Scoring table: 1 gcgcagcagtgatcgcgcagcatl 25

Searched: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Word size : 9

Total number of hits satisfying chosen parameters: 12796

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: listing first 100 summaries

Database :

1: Issued Patents_NA: *
2: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/Backfill1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	% Match	Query Length	ID	Description
1	25	100.0	1041 2 US-08-401-068-7	Sequence 7, Appli
2	25	100.0	1041 2 US-08-846-338-7	Sequence 7, Appli
3	25	100.0	5872 3 US-08-411-768B-1	Sequence 1, Appli
4	25	100.0	5872 3 US-08-411-768B-6	Sequence 6, Appli
5	16	64.0	4411529 4 US-09-103-840A-1	Sequence 1, Appli
6	14	56.0	718 4 US-09-142-078-45	Sequence 45, Appli
7	14	56.0	718 4 US-09-357-141-45	Sequence 45, Appli
8	14	56.0	2193 1 US-08-420-235B-16	Sequence 16, Appli
9	14	56.0	2193 3 US-08-793-624-16	Sequence 16, Appli
10	14	56.0	2193 3 PCT-US95-10194-16	Sequence 16, Appli
11	14	56.0	2973 2 US-08-343-101A-5	Sequence 5, Appli
12	14	56.0	2973 2 US-09-183-688-5	Sequence 5, Appli
13	14	56.0	4182 1 US-07-973-257-1	Sequence 1, Appli
14	14	56.0	5362 2 US-08-853-310-3	Sequence 3, Appli
15	14	56.0	5362 2 US-08-420-235B-1	Sequence 1, Appli
16	14	56.0	20710 1 US-08-793-624-1	Sequence 1, Appli
17	14	56.0	20710 5 PCT-US95-10194-1	Sequence 1, Appli
18	14	56.0	35100 2 US-08-770-379-18	Sequence 18, Appli
19	14	56.0	35100 2 US-08-757-669A-18	Sequence 18, Appli
20	14	56.0	304 4 US-09-060-756-726	Sequence 726, App
21	13	52.0	343 4 US-09-060-756-570	Sequence 570, App
22	13	52.0	411 2 US-08-640-978C-2	Sequence 2, Appli
23	13	52.0	414 2 US-08-640-978C-4	Sequence 4, Appli
24	13	52.0	447 4 US-08-836-075A-51	Sequence 51, Appli
25	13	52.0	476 4 US-08-998-416-978	Sequence 978, App
26	13	52.0	604 4 US-09-342-653-3	Sequence 3, Appli
27	13	52.0	1128 1 US-07-872-678A-11	Sequence 11, Appli

28	13	52.0	1202 2 US-08-428-414A-1	Sequence 1, Appli
29	13	52.0	1260 4 US-08-943-731-169	Sequence 169, App
30	13	52.0	1365 4 US-09-319-892-1	Sequence 1, Appli
31	13	52.0	1624 2 US-07-862-588B-6	Sequence 6, Appli
32	13	52.0	1775 2 US-07-862-588B-5	Sequence 5, Appli
33	13	52.0	1875 2 US-09-258-373-21	Sequence 21, Appli
34	13	52.0	1981 3 US-09-017-706-3	Sequence 3, Appli
35	13	52.0	1981 3 US-09-017-706-4	Sequence 4, Appli
36	13	52.0	1981 3 US-09-017-706-5	Sequence 5, Appli
37	13	52.0	1981 3 US-09-017-706-6	Sequence 6, Appli
38	13	52.0	1981 3 US-09-017-706-7	Sequence 7, Appli
39	13	52.0	1981 3 US-09-017-706-8	Sequence 8, Appli
40	13	52.0	2223 1 US-08-439-818A-8	Sequence 8, Appli
41	13	52.0	2223 1 US-08-439-818A-8	Sequence 8, Appli
42	13	52.0	2223 2 US-08-751-965-8	Sequence 8, Appli
43	13	52.0	2223 2 US-08-738-975-8	Sequence 8, Appli
44	13	52.0	2223 2 US-08-728-626-8	Sequence 8, Appli
45	13	52.0	2223 3 US-08-808-599A-8	Sequence 8, Appli
46	13	52.0	2290 1 US-08-584-226-1	Sequence 1, Appli
47	13	52.0	2763 3 US-08-941-445A-16	Sequence 16, Appli
48	13	52.0	3618 1 US-07-872-678A-36	Sequence 36, Appli
49	13	52.0	3670 1 US-08-386-09472-16	Sequence 3, Appli
50	13	52.0	3670 2 US-08-600-452A-3	Sequence 3, Appli
51	13	52.0	4223 4 US-09-541-782-1	Sequence 5, Appli
52	13	52.0	4615 1 US-08-188-582-1	Sequence 1, Appli
53	13	52.0	4615 1 US-08-646-715-1	Sequence 1, Appli
54	13	52.0	5319 1 US-08-169-927-1	Sequence 1, Appli
55	13	52.0	8878 1 US-08-759-444-2	Sequence 2, Appli
56	13	52.0	8982 3 US-08-976-255-5	Sequence 5, Appli
57	13	52.0	9880 3 US-08-680-897-1	Sequence 1, Appli
58	13	52.0	17606 4 US-08-943-731-4	Sequence 4, Appli
59	13	52.0	50341 2 US-08-247-901C-1	Sequence 1, Appli
60	13	52.0	50341 2 US-09-075-904-1	Sequence 1, Appli
61	13	52.0	52297 4 US-09-426-436-1	Sequence 1, Appli
62	13	52.0	52297 4 US-08-705-557-1	Sequence 1, Appli
63	13	52.0	4433765 4 US-09-103-840A-2	Sequence 2, Appli
64	13	52.0	4411529 4 US-09-103-840A-1	Sequence 1, Appli
65	12	48.0	19 3 US-08-851-843A-92	Sequence 92, Appli
66	12	48.0	19 4 US-08-974-549A-384	Sequence 384, App
67	12	48.0	19 4 US-08-854-050-92	Sequence 92, Appli
68	12	48.0	69 2 US-08-488-402A-16	Sequence 16, Appli
69	12	48.0	69 2 US-08-484-552A-16	Sequence 16, Appli
70	12	48.0	69 5 PCT-US96-09472-16	Sequence 4, Appli
71	12	48.0	372 3 US-07-687-819-4	Sequence 4, Appli
72	12	48.0	386 4 US-08-998-416-93	Sequence 4, Appli
73	12	48.0	390 3 US-07-687-819-5	Sequence 5, Appli
74	12	48.0	393 3 US-08-925-230-5	Sequence 5, Appli
75	12	48.0	393 3 US-08-925-230-6	Sequence 6, Appli
76	12	48.0	438 2 US-08-743-637B-184	Sequence 184, App
77	12	48.0	484 6 US-08-413-803-28	Sequence 28, Appli
78	12	48.0	498 1 US-07-749-446-1	Sequence 1, Appli
79	12	48.0	579 1 US-07-149-101A-5	Sequence 5, Appli
80	12	48.0	603 4 PCT-US94-12873-5	Sequence 1063, Ap
81	12	48.0	615 4 US-08-998-416-1063	Sequence 56, Appli
82	12	48.0	656 4 US-08-998-416-56	Sequence 1, Appli
83	12	48.0	723 2 US-08-618-911-1	Sequence 237, App
84	12	48.0	754 4 US-08-998-416-237	Sequence 11, Appli
85	12	48.0	780 4 US-08-855-825-11	Sequence 11, Appli
86	12	48.0	780 4 US-08-855-825-13	Sequence 11, Appli
87	12	48.0	782 1 US-07-865-878A-1	Sequence 1, Appli
88	12	48.0	782 2 US-08-308-736A-20	Sequence 20, Appli
89	12	48.0	782 2 US-08-603-010-1	Sequence 1, Appli
90	12	48.0	792 2 US-09-154-083-23	Sequence 23, Appli
91	12	48.0	841 3 US-08-990-823-39	Sequence 39, Appli
92	12	48.0	842 3 US-08-983-409-1	Sequence 1, Appli
93	12	48.0	946 2 US-08-408-095-26	Sequence 26, Appli
94	12	48.0	954 1 US-08-624-125-9	Sequence 9, Appli
95	12	48.0	1029 3 US-08-911-855-1	Sequence 1, Appli
96	12	48.0	1029 3 US-09-479-409-1	Sequence 17, Appli
97	12	48.0	1029 4 US-08-459-586-17	Sequence 17, Appli
98	12	48.0	1174 2 US-08-282-698-17	Sequence 17, Appli
99	12	48.0	1174 2 US-08-861-774E-85	Sequence 85, Appli
100	12	48.0	1200 4	

ALIGNMENTS

```
RESULT 1
; US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: Linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "Biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcacagtgatcgccagcatt 25
DB 79 ggcgcacagtgatcgccagcatt 103

RESULT 2
; US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5859719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
```

```
ADDRESS: NO. 5869719artis Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: Linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "Biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcacagtgatcgccagcatt 25
DB 79 ggcgcacagtgatcgccagcatt 103

RESULT 3
; US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect
; SOFTWARE: Version 5.1
; CURRENT APPLICATION DATA:
```



```

APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=117
OTHER INFORMATION: /product="Biotin synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="biob"
OTHER INFORMATION: /number=1
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=2295
OTHER INFORMATION: /function="Involved in pimeloyl-CoA synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /gene="bioc"
OTHER INFORMATION: /number=3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /EC_number=2.6.1.62
OTHER INFORMATION: /product="DAPA synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="biob"
OTHER INFORMATION: /number=5
OTHER INFORMATION: /standard_name="S-adenosyl-L-methionine:8-amino-7-oxononanoate"
OTHER INFORMATION: /aminoacids="aminoacids"
FEATURE:
NAME/KEY: CDS
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=5098
OTHER INFORMATION: /function="unknown, involved in biotin synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="ORF1"
OTHER INFORMATION: /number=6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="promoter ptac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name="promoter ptac"

```

```

FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name="bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name="biob RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name="ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name="rho-independent"
OTHER INFORMATION: /transcriptional_terminator="rho-independent"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function="promoter ptac"
OTHER INFORMATION: /evidence=EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

```

```

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 6.2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 ggcgcagcagtgatcgcgcagcatt 25
Db 195 GCGCAGCAGCTGCATCGCCGCACTT 219

```

```

RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect
; SOFTWARE: Version 5.1

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: PB030A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /EC_number= 1154
OTHER INFORMATION: /product= "KAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioF"
OTHER INFORMATION: /number= 2
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /EC_number= 3043
OTHER INFORMATION: /product= "DPR synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioD"
OTHER INFORMATION: /number= 4
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name= "bioF RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name= "bioD RBS"
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6

Query Match          100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 6; 2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcagcaggtgcatcgccacatt 25
|||||
Db 195 GCGCAGCAGGTGTCATCGCCACATT 219

RESULT 5
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
```

```

GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-2007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match          64.0%; Score 16; DB 4; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 2; 6;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cgcagcaggtgcatcg 17
|||||
Db 4254432 cgcagcaggtgcatcg 4254447

RESULT 6
US-09-142-078-45/C
; Sequence 45, Application US/09142078
; Patent No. 6172041
GENERAL INFORMATION:
APPLICANT: McCabe, R. Tyler
APPLICANT: Zhou, Li-Ming
APPLICANT: Layer, Richard T.
TITLE OF INVENTION: Use of Conantokins
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
STREET: 555 Thirteenth Street, N.W., Suite 701-E
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/142,078
FILING DATE: 10-FEB-1999
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: WO US97/12652
FILING DATE: 21-JUL-1997
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/762,377
FILING DATE: 06-DEC-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/684,750
FILING DATE: 22-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 2314-135.A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-783-6040
TELEFAX: 202-783-6031
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 base pairs
```

```

: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (CDNA)
: ORIGINAL SOURCE:
: ORGANISM: Conus geographus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 110..409
: US-09-142-078-45

```

```

Query Match          56.0%; Score 14; DB 4; Length 718;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 7 caggtgcatgcgcca 20
    |||
DB 118 CAGGTGATCGCCA 105

```

```

RESULT 7
US-09-357-141-45/c
: Sequence 45, Application US/09357141
: Patent No. 6277825
: GENERAL INFORMATION:
: APPLICANT: Olivera, Baldomero M.
: APPLICANT: McIntosh, J. Michael
: APPLICANT: McCabe, R. Tyler
: APPLICANT: Layer, Richard T.
: APPLICANT: Zhou, Li-Ming
: TITLE OF INVENTION: Use of Conantokins for Treating Pain
: FILE REFERENCE: 2314-171
: CURRENT APPLICATION NUMBER: US/09/357,141
: CURRENT FILING DATE: 1999-07-20
: PRIOR APPLICATION NUMBER: US 09/283,277
: PRIOR FILING DATE: 1999-04-01
: PRIOR APPLICATION NUMBER: US 09/142,078
: PRIOR FILING DATE: 1999-02-10
: PRIOR APPLICATION NUMBER: WO US97/12652
: PRIOR FILING DATE: 1997-07-21
: PRIOR FILING DATE: 1996-12-06
: PRIOR FILING DATE: 1996-07-22
: PRIOR FILING DATE: 1996-07-22
: NUMBER OF SEQ ID NOS: 71
: SOFTWARE: PatentIn Ver. 2.0
: SEO ID NO 45
: LENGTH: 718
: TYPE: DNA
: ORGANISM: Conus geographus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (110)..(409)
: US-09-357-141-45

Query Match          56.0%; Score 14; DB 4; Length 718;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 caggtgcatgcgcca 20
    |||
DB 118 CAGGTGATCGCCA 105

RESULT 8
US-08-420-235B-16/c
: Sequence 16, Application US/08420235B
: Patent No. 5801042
: GENERAL INFORMATION:
: APPLICANT: Chang, Yuan
: APPLICANT: Moore, Patrick S.

```

```

: TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
: NUMBER OF SEQUENCES: 47
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cooper & Dunham LLP
: STREET: 1185 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/420,235B
: FILING DATE:
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 45185-B
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 278-0400
: TELEFAX: (212) 391-0525
: INFORMATION FOR SEQ ID NO: 16:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2193 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEetical: N
: ANTI-SENSE: N
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..2193
: OTHER INFORMATION:
: US-08-420-235B-16

```

```

Query Match          56.0%; Score 14; DB 1; Length 2193;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgccaqcat 24
    |||
DB 46 TGCATCGCCAGCAT 33

```

```

RESULT 9
US-08-793-624-16/c
: Sequence 16, Application US/08793624C
: Patent No. 6150093
: GENERAL INFORMATION:
: APPLICANT: Chang, Yuan
: APPLICANT: Moore, Patrick S.
: TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And
: TITLE OF INVENTION: Uses Thereof
: FILE REFERENCE: 45185-C-PCT-US/JPW
: CURRENT APPLICATION NUMBER: US/08/793,624C
: CURRENT FILING DATE: 1997-02-18
: NUMBER OF SEQ ID NOS: 58
: SOFTWARE: PatentIn Ver. 2.1
: SEO ID NO 16
: LENGTH: 2193
: TYPE: DNA
: ORGANISM: Kaposi's sarcoma-associated herpesvirus
: US-08-793-624-16

Query Match          56.0%; Score 14; DB 3; Length 2193;

```

Best Local Similarity 100.0%; Pred. No. 31;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgcacgat 24
|||||
Db 46 tgcacgcacgat 33

RESULT 10

PCT-US95-10194-16/c
; Sequence 16, Application PC/TUS9510194
; GENERAL INFORMATION:
; APPLICANT: The Trustees of Columbia University in the City of New York
; APPLICANT: City
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10194
FILING DATE:

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-C-PCT/JWP/MS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:
LENGTH: 2193 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2193
OTHER INFORMATION:
PCT-US95-10194-16

Query Match 56.0%; Score 14; DB 5; Length 2193;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgcacgat 24
|||||
Db 46 tgcacgcacgat 33

RESULT 11

US-08-343-101A-5/c
; Sequence 5, Application US/08343101A
; Patent No. 5830759
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma

TITLE OF INVENTION: Virus Sequences And Uses Thereof
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,101A
FILING DATE:

CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: White Esq., John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0526
INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:
LENGTH: 2973 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
US-08-343-101A-5

Query Match 56.0%; Score 14; DB 2; Length 2973;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgcacgat 24
|||||
Db 2182 tgcacgcacgat 2169

RESULT 12

US-09-183-688-5/c
; Sequence 5, Application US/09183688
; Patent No. 6093550
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.

TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma
TITLE OF INVENTION: Virus Sequences And Uses Thereof
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/183,688
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/343,101
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: White Esq., John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0526
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2973 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
US-09-183-688-5

Query Match 56.0%; Score 14; DB 3; Length 2973;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 tgcacgcacgacat 24
|||||
DB 2182 tgcacgcacgacat 2169

RESULT 13
US-07-973-257-1/c
Sequence 1, Application US/07973257
Patent No. 5378820
GENERAL INFORMATION:
APPLICANT: Keeler, Jr, Calvin L.
APPLICANT: Dohms, John E.
TITLE OF INVENTION: Gene Encoding Cytohesin
TITLE OF INVENTION: Protein of Mycoplasma Gallisepticum and Its Use
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Connolly and Hutz
STREET: 1220 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: U.S.A.
ZIP: 19899
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM/PC or Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/973,257
FILING DATE: 19921109
CLASSIFICATION: 435
PRIOR APPLICATION DATA: No. 5378820e
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4182 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double stranded
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mycoplasma Gallisepticum
STRAIN: S6
US-07-973-257-1

Query Match 56.0%; Score 14; DB 1; Length 4182;

Best Local Similarity 100.0%; Pred. No. 31;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 gcacgacggtgcac 16
|||||
DB 1260 gcacgacggtgcac 1247

RESULT 14
US-08-853-310-3/c
Sequence 3, Application US/08853310
Patent No. 5948640
GENERAL INFORMATION:
APPLICANT: Randazzo, Filippo
TITLE OF INVENTION: Mammalian Additional Sex Combs (Asx) Acts as a Tumor Suppre
CORRESPONDENCE ADDRESS: 4
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: U.S.A.
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/853,310
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Guth, Joseph H.
REGISTRATION NUMBER: 31,261
REFERENCE/DOCKET NUMBER: 1228, 003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 923-3888
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5362 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-853-310-3

Query Match 56.0%; Score 14; DB 2; Length 5362;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 gcacgacggtgcac 16
|||||
DB 3354 gcacgacggtgcac 3341

RESULT 15
US-08-420-235B-1/c
Sequence 1, Application US/08420235B
Patent No. 5801042
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York

COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/420,235B
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 20710 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
US-08-420-235B-1

Query Match 56.0%; Score 14; DB 1; Length 20710;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgccacgat 24
|||||
Db 2138 TGCATGCCACGAT 2125

RESULT 16
US-08-793-624-1/C
Sequence 1, Application US/08793624C
Patent No. 6150093
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And
TITLE OF INVENTION: Uses Thereof
FILE REFERENCE: 45185-C-PCT-US/JPM
CURRENT APPLICATION NUMBER: US/08/793,624C
CURRENT FILING DATE: 1997-02-18
NUMBER OF SEQ ID NOS: 58
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 20710
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-08-793-624-1

Query Match 56.0%; Score 14; DB 3; Length 20710;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgccacgat 24
|||||
Db 2138 TGCATGCCACGAT 2125

RESULT 17
PCT-US95-10194-1/C
Sequence 1, Application PC/TUS9510194
GENERAL INFORMATION:

APPLICANT: The Trustees of Columbia University in the City of New York
APPLICANT: City
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10194
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPM/MS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 20710 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
PCT-US95-10194-1

Query Match 56.0%; Score 14; DB 5; Length 20710;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgccacgat 24
|||||
Db 2138 TGCATGCCACGAT 2125

RESULT 18
US-08-770-379-18/C
Sequence 18, Application US/08770379
Patent No. 5849564
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,379
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 35100 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-770-379-18

Query Match
Best Local Similarity 56.0%; Score 14; DB 2; Length 35100;
Pred. No. 30;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 tgcacgcgcacat 24
|||||
Db 2058 tgcacgcgcacat 2045

RESULT 19
US-08-757-669A-18/c
Sequence 18, Application US/08757669A
Patent No. 6183751
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: UNTOQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/757,669A
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-F
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 35100 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-08-757-669A-18

Query Match
Best Local Similarity 56.0%; Score 14; DB 4; Length 35100;
Pred. No. 30;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 tgcacgcgcacat 24
|||||
Db 2058 tgcacgcgcacat 2045

RESULT 20
US-09-060-756-726
Sequence 726, Application US/09060756
Patent No. 6183957
GENERAL INFORMATION:
APPLICANT: Buchrieser-Brosch, Roland
APPLICANT: Gordon, Stephen
APPLICANT: Billault, Alain
TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
FILE REFERENCE: 3495-0169
CURRENT APPLICATION NUMBER: US/09/060,756
CURRENT FILING DATE: 1998-04-16
NUMBER OF SEQ ID NOS: 743
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 570
LENGTH: 343
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
US-09-060-756-570

Query Match
Best Local Similarity 52.0%; Score 13; DB 4; Length 304;
Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 gtgcacgcgcacg 22
|||||
Db 136 gtgcacgcgcacg 148

RESULT 21
US-09-060-756-570
Sequence 570, Application US/09060756
Patent No. 6183957
GENERAL INFORMATION:
APPLICANT: Cole, Stewart
APPLICANT: Buchrieser-Brosch, Roland
APPLICANT: Gordon, Stephen
APPLICANT: Billault, Alain
TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
FILE REFERENCE: 3495-0169
CURRENT APPLICATION NUMBER: US/09/060,756
CURRENT FILING DATE: 1998-04-16
NUMBER OF SEQ ID NOS: 743
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 570
LENGTH: 343
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
US-09-060-756-570

Query Match 52.0%; Score 13; DB 4; Length 343;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 ggcacgcgcgcgc 22
|||||
DB 135 ggcacgcgcgcgc 147

RESULT 22

US-08-640-978C-2/c
; Sequence 2, Application US/08640978C
; Patent No. 5932781
; GENERAL INFORMATION:
; APPLICANT: TAKANO, MITSUO
; APPLICANT: ONO, HISATO
; APPLICANT: YAMADA, HIROYUKI
; APPLICANT: YAMATOYA, KAZUHIKO
; TITLE OF INVENTION: ECTOINE SYNTHASE GENE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/640,978C
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C J
; REGISTRATION NUMBER: 32,350
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-640-978C-2

Query Match 52.0%; Score 13; DB 2; Length 411;
Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcacgcgcgcgc 13
|||||
DB 294 ggcgcacgcgcgcgc 282

RESULT 23

US-08-640-978C-4/c
; Sequence 4, Application US/08640978C
; Patent No. 5932781
; GENERAL INFORMATION:
; APPLICANT: TAKANO, MITSUO
; APPLICANT: ONO, HISATO
; APPLICANT: YAMADA, HIROYUKI
; APPLICANT: YAMATOYA, KAZUHIKO
; TITLE OF INVENTION: ECTOINE SYNTHASE GENE
; NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
STREET: PO BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/640,978C
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: FARACI, C J
REGISTRATION NUMBER: 32,350
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 414 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..411
US-08-640-978C-4

Query Match 52.0%; Score 13; DB 2; Length 414;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcacgcgcgcgc 13
|||||
DB 294 ggcgcacgcgcgcgc 282

RESULT 24

US-08-636-075A-51
; Sequence 51, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,075A
; FILING DATE: 21 Apr 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04155
; FILING DATE: 23 Oct 1995


```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-836-075A-51

Query Match
; Score 13; DB 4; Length 447;
; Best Local Similarity 100.0%; Pred. No. 1e+02;
; Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 catgcagcagcatt 25
DB 406 CATGCCGACGACATT 418

RESULT 25
US-08-998-416-978/c
; Sequence 978, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippson, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jürgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPHII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NO. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 978:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 476 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1602UP
; US-08-998-416-978

Query Match
; Score 13; DB 4; Length 476;
; Best Local Similarity 100.0%; Pred. No. 1e+02;
; Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgc 13
DB 328 GCCGACGACGAGTGC 316

RESULT 26
US-09-342-653-3
; Sequence 3, Application US/09342653
; Patent No. 6306632
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Chromatin Associated Proteins
; FILE REFERENCE: BB-1118
; CURRENT APPLICATION NUMBER: US/09/342,653
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: 60/092,841
; EARLIER FILING DATE: July 14, 1998
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 3
; LENGTH: 604
; TYPE: DNA
; ORGANISM: Oryza sativa
; US-09-342-653-3

Query Match
; Score 13; DB 4; Length 604;
; Best Local Similarity 100.0%; Pred. No. 1e+02;
; Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 agcagtgatcatcg 17
DB 458 agcagtgatcatcg 470

RESULT 27
US-07-872-678A-11
; Sequence 11, Application US/07872678A
; Patent No. 5541060
; GENERAL INFORMATION:
; APPLICANT: Bell, Graeme, et al.
; TITLE OF INVENTION: DETECTION OF EARLY-ONSET
; TITLE OF INVENTION: NON-INSULIN-DEPENDENT DIABETES MELLITUS
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: Post Office Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/07/872,678A
FILING DATE: 22-APRIL-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coughlin, Daniel F.
REGISTRATION NUMBER: 36,111
REFERENCE/DOCKET NUMBER: ARCD016
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-787-1400
TELEFAX: 713-789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1128 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-872-678A-11

Query Match 52.0%; Score 13; DB 1; Length 1128;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 cgcagcaggtgca 14
|||||
Db 904 CGCAGCAGGTGCA 916

RESULT 28
US-08-428-414A-1
Sequence 1, Application US/08428414A
Patent No. 5912166
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TITLE OF INVENTION: LEISHMANIASIS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428,414A
FILING DATE: 21-APR-1995
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Kadlecik, Ann F.
REGISTRATION NUMBER: 39,244
REFERENCE/DOCKET NUMBER: 210121,407
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1202 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 30..998
US-08-428-414A-1

Query Match 52.0%; Score 13; DB 2; Length 1202;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 cgcagcaggtgca 14
|||||
Db 133 CGCAGCAGGTGCA 145

RESULT 29
US-08-943-731-169
Sequence 169, Application US/08943731
Patent No. 6265157
GENERAL INFORMATION:
APPLICANT: PROCKOP, DARWIN J.
APPLICANT: SPOTILA, LORETTA D.
APPLICANT: DELTAS, CONSTANTINOS D.
APPLICANT: SEREDA, LARISA
APPLICANT: LARSON, ANDREA W.
APPLICANT: PACK, MICHAEL
APPLICANT: COLIGE, ALAIN
APPLICANT: EARLY, JAMES
APPLICANT: KORRKO, JARMO
APPLICANT: ALA-KORRKO, LEENA, et al.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
NUMBER OF SEQUENCES: 666
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANITCH SCHWARZ JACOBS & NADEL, P.C.
STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
STREET: FLR.
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA
ZIP: 19103-7086
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,731
FILING DATE: 03-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,322
FILING DATE: 14-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/803,628
FILING DATE: 03-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: DOYLE LEARY Ph.D., KATHRYN
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 9598-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1284
TELEFAX: 215-567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 169:
SEQUENCE CHARACTERISTICS:
LENGTH: 1260 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-943-731-169

Query Match 52.0%; Score 13; DB 4; Length 1260;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 gccagcagtgcat 15
|||||
Db 1199 GCACGAGTGTCAT 1211

RESULT 30
US-09-319-892-1
Sequence 1, Application US/09319892
Patent No. 6177616
GENERAL INFORMATION:
APPLICANT: BARTSCH, Klaus
APPLICANT: KRIETE, Guido
APPLICANT: BROER, Inge
APPLICANT: PUHLER, Alfred
TITLE OF INVENTION: NOVEL GENES CODING FOR AMINO ACID DEACETYLASES WITH
TITLE OF INVENTION: SPECIFICITY FOR N-ACETYL-L-PHOSPHINOETHYLICIN, THEIR
TITLE OF INVENTION: ISOLATION AND THEIR USE
FILE REFERENCE: 514412-2005
CURRENT APPLICATION NUMBER: US/09/319,892
EARLIER FILING DATE: 1999-06-14
EARLIER APPLICATION NUMBER: PCT/EP97/06755
EARLIER FILING DATE: 1997-12-03
EARLIER APPLICATION NUMBER: 19652284.6
EARLIER FILING DATE: 1996-12-16
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1365
TYPE: DNA
ORGANISM: Stenotrophomonas maltophilia
US-09-319-892-1

Query Match 52.0%; Score 13; DB 4; Length 1365;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gccagcagtgatc 13
|||||
Db 1139 gccagcagtgatc 1151

RESULT 31
US-07-862-588B-6
Sequence 6, Application US/07862588B
Patent No. 5916796
GENERAL INFORMATION:
APPLICANT: Joergensen, Per Linnae
APPLICANT: Sch lein, Martin
APPLICANT: Hansen, Christian
TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESS: 405 Lexington Avenue, 62nd floor
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10017
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/862,588B
FILING DATE: 19920727
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 164/90
FILING DATE: 19-JAN-1990
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/DK91/00013
FILING DATE: 18-JAN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Zelson, Steve T. / Lambiris, Elias J.
REGISTRATION NUMBER: 30,335 / 33,728
REFERENCE/DOCKET NUMBER: 3425, 204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1624 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Bacillus lautus
STRAIN: NCIMB 40250
FEATURE:
NAME/KEY: CDS
LOCATION: 30..1607
OTHER INFORMATION:
US-07-862-588B-6

Query Match 52.0%; Score 13; DB 2; Length 1624;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 caagtgatcgc 19
|||||
Db 934 CAGTGATCGCC 946

RESULT 32
US-07-862-588B-5
Sequence 5, Application US/07862588B
Patent No. 5916796
GENERAL INFORMATION:
APPLICANT: Joergensen, Per Linnae
APPLICANT: Sch lein, Martin
APPLICANT: Hansen, Christian
TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESS: 405 Lexington Avenue, 62nd floor
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10017
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/862,588B
FILING DATE: 19920727
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 164/90
FILING DATE: 19-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK91/00013
FILING DATE: 18-JAN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Zelson, Steve T. / Lambiris, Elias J.
REGISTRATION NUMBER: 30,335 / 33,728
REFERENCE/DOCKET NUMBER: 3425, 204-US
TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1775 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 30..(1625,1775)
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /partial
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /transl_except= (pos: 1446 .. 1458, aa:
; OTHER INFORMATION: (OTR)
US-07-862-588B-5

```

```

Query Match          52.0%; Score 13; DB 3; Length 1775;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 7 cgcagcagatgcc 19
    |||||
Db 934 CAGGTGCATGCC 946

```

```

RESULT 33
US-09-258-373-21/c
; Sequence 21, Application US/09258373
; Patent No. 6150110
; GENERAL INFORMATION:
; APPLICANT: Fletcher, Jonathan A.
; TITLE OF INVENTION: HMGI(Y)-LAMA4* FUSION ONCOGENE,
; FILE REFERENCE: B0801/7135/ERP
; CURRENT APPLICATION NUMBER: US/09/258,373
; EARLIER FILING DATE: 1999-02-26
; EARLIER APPLICATION NUMBER: 60/076,401
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 1875
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-258-373-21

```

```

Query Match          52.0%; Score 13; DB 3; Length 1875;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 cgcagcagatgca 14
    |||||
Db 838 CGCAGCAGTGCA 826

```

```

RESULT 34
US-09-017-706-3
; Sequence 3, Application US/09017706A
; Patent No. 6087147
; GENERAL INFORMATION:
; APPLICANT: ITO, YOSHIFUMI
; TITLE OF INVENTION: A-AMYLASE GENE HAVING ABILITY FOR HIGHLY PRODUCING
; TITLE OF INVENTION: MALTOTENAOSE, VECTOR CONTAINING SAID GENE AND

```

```

; TITLE OF INVENTION: TRANSFORMANT
; FILE REFERENCE: 8361-0003-0
; CURRENT APPLICATION NUMBER: US/09/017,706A
; CURRENT FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: JP 305071/1997
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1981
; TYPE: DNA
; ORGANISM: Pseudomonas sp., Strain KO-8940
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(1848)
; NAME/KEY: mat_peptide
; LOCATION: (85)..(1848)
; OTHER INFORMATION: METHOD OF DETERMINING CHARACTERISTIC: E
; FEATURE:
; OTHER INFORMATION: MUTATED GENOMIC DNA
; FEATURE:
; OTHER INFORMATION: PLASMID POS3410F57
US-09-017-706-3

```

```

Query Match          52.0%; Score 13; DB 3; Length 1981;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 cgcagcagatgca 14
    |||||
Db 1530 cgcagcagatgca 1542

```

```

RESULT 35
US-09-017-706-4
; Sequence 4, Application US/09017706A
; Patent No. 6087147
; GENERAL INFORMATION:
; APPLICANT: ITO, YOSHIFUMI
; TITLE OF INVENTION: A-AMYLASE GENE HAVING ABILITY FOR HIGHLY PRODUCING
; TITLE OF INVENTION: MALTOTENAOSE, VECTOR CONTAINING SAID GENE AND
; FILE REFERENCE: 8361-0003-0
; CURRENT APPLICATION NUMBER: US/09/017,706A
; CURRENT FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: JP 305071/1997
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1981
; TYPE: DNA
; ORGANISM: Pseudomonas sp., Strain KO-8940
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(1848)
; NAME/KEY: mat_peptide
; LOCATION: (85)..(1848)
; OTHER INFORMATION: METHOD OF DETERMINING CHARACTERISTIC: E
; FEATURE:
; OTHER INFORMATION: MUTATED GENOMIC DNA
; OTHER INFORMATION: PLASMID POS3410H57
US-09-017-706-4

```

```

Query Match          52.0%; Score 13; DB 3; Length 1981;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 2 cgcagcaggtgca 14
|||||
Db 1530 cgcagcaggtgca 1542

RESULT 36

US-09-017-706-5
; Sequence 5, Application US/09017706A
; Patent No. 6087147
; GENERAL INFORMATION:
; APPLICANT: ITO, YOSHIFUMI
; TITLE OF INVENTION: A-AMYLASE GENE HAVING ABILITY FOR HIGHLY PRODUCING
; TITLE OF INVENTION: MALTOPEPTASE, VECTOR CONTAINING SAID GENE AND
; FILE REFERENCE: 8361-0003-0
; CURRENT APPLICATION NUMBER: US/09/017,706A
; EARLIER FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: JP 305071/1997
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1981
; TYPE: DNA
; ORGANISM: Pseudomonas sp., Strain KO-8940
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(1848)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (85)..(1848)
; OTHER INFORMATION: METHOD OF DETERMINING CHARACTERISTIC: E
; FEATURE:
; OTHER INFORMATION: MUTATED GENOMIC DNA
; FEATURE:
; OTHER INFORMATION: PLASMID POS34100L57
US-09-017-706-5

Query Match 52.0%; Score 13; DB 3; Length 1981;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cgcagcaggtgca 14
|||||
Db 1530 cgcagcaggtgca 1542

RESULT 37

US-09-017-706-6
; Sequence 6, Application US/09017706A
; Patent No. 6087147
; GENERAL INFORMATION:
; APPLICANT: ITO, YOSHIFUMI
; TITLE OF INVENTION: A-AMYLASE GENE HAVING ABILITY FOR HIGHLY PRODUCING
; TITLE OF INVENTION: MALTOPEPTASE, VECTOR CONTAINING SAID GENE AND
; FILE REFERENCE: 8361-0003-0
; CURRENT APPLICATION NUMBER: US/09/017,706A
; CURRENT FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: JP 305071/1997
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1981
; TYPE: DNA
; ORGANISM: Pseudomonas sp., Strain KO-8940
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(1848)
; FEATURE:
; NAME/KEY: mat_peptide

; LOCATION: (85)..(1848)
; FEATURE:
; OTHER INFORMATION: MUTATED GENOMIC DNA
; FEATURE:
; OTHER INFORMATION: PLASMID POS3410F139
US-09-017-706-6

Query Match 52.0%; Score 13; DB 3; Length 1981;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cgcagcaggtgca 14
|||||
Db 1530 cgcagcaggtgca 1542

RESULT 38

US-09-017-706-7
; Sequence 7, Application US/09017706A
; Patent No. 6087147
; GENERAL INFORMATION:
; APPLICANT: ITO, YOSHIFUMI
; TITLE OF INVENTION: A-AMYLASE GENE HAVING ABILITY FOR HIGHLY PRODUCING
; TITLE OF INVENTION: MALTOPEPTASE, VECTOR CONTAINING SAID GENE AND
; FILE REFERENCE: 8361-0003-0
; CURRENT APPLICATION NUMBER: US/09/017,706A
; CURRENT FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: JP 305071/1997
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1981
; TYPE: DNA
; ORGANISM: Pseudomonas sp., Strain KO-8940
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(1848)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (85)..(1848)
; OTHER INFORMATION: METHOD OF DETERMINING CHARACTERISTIC: E
; FEATURE:
; OTHER INFORMATION: MUTATED GENOMIC DNA
; FEATURE:
; OTHER INFORMATION: PLASMID POS3410H139
US-09-017-706-7

Query Match 52.0%; Score 13; DB 3; Length 1981;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cgcagcaggtgca 14
|||||
Db 1530 cgcagcaggtgca 1542

RESULT 39

US-09-017-706-8
; Sequence 8, Application US/09017706A
; Patent No. 6087147
; GENERAL INFORMATION:
; APPLICANT: ITO, YOSHIFUMI
; TITLE OF INVENTION: A-AMYLASE GENE HAVING ABILITY FOR HIGHLY PRODUCING
; TITLE OF INVENTION: MALTOPEPTASE, VECTOR CONTAINING SAID GENE AND
; FILE REFERENCE: 8361-0003-0
; CURRENT APPLICATION NUMBER: US/09/017,706A
; CURRENT FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: JP 305071/1997

EARLIER FILING DATE: 1997-10-21
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO: 8
LENGTH: 1981
TYPE: DNA
ORGANISM: Pseudomonas sp., Strain KO-8940
FEATURE:
NAME/KEY: CDS
LOCATION: (7)..(1848)
FEATURE:
NAME/KEY: mat.peptide
LOCATION: (85)..(1848)
FEATURE:
OTHER INFORMATION: MUTATED GENOMIC DNA
FEATURE:
OTHER INFORMATION: PLASMID: POS3410L139
US-09-017-706-8

Query Match 52.0%; Score 13; DB 3; Length 1981;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cgcagcaggtgc 14
|||||
DB 1530 cgcagcaggtgc 1542

RESULT 40
US-08-317-522A-8/c
Sequence 8, Application US/08317522A
Patent No. 5599918
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Trophinin and Trophinin-Assisting
NUMBER OF INVENTION: Proteins
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,522A
FILING DATE: 04-OCT-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 9991
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2223 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 199..2223
US-08-317-522A-8

Query Match 52.0%; Score 13; DB 1; Length 2223;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcagcaggtgc 13
|||||
DB 2118 GCGCAGCAGGTGC 2106

RESULT 41
US-08-439-818A-8/c
Sequence 8, Application US/08439818A
Patent No. 5654145
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Trophinin and Trophinin-Assisting
NUMBER OF INVENTION: Proteins
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,818A
FILING DATE: 12-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/317,522
FILING DATE: 04-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 1563
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2223 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 199..2223
US-08-439-818A-8

Query Match 52.0%; Score 13; DB 1; Length 2223;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcagcaggtgc 13
|||||
DB 2118 GCGCAGCAGGTGC 2106

RESULT 42
US-08-751-965-8/c
Sequence 8, Application US/08751965
Patent No. 5858360
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Trophinin and Trophinin-Assisting
Proteins

```

;
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,965
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/439,818
; FILING DATE: 12-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2223 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 199..2223
;
; US-08-751-965-8
;
;
; Query Match 52.0%; Score 13; DB 2; Length 2223;
; Best Local Similarity 100.0%; Pred. NO. 1e+02;
; Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 gccgcagcaggtgc 13
; |
; Db 2118 GCCGACGACGTGC 2106
;
;
; RESULT 43
; US-08-738-975-8/C
; Sequence 8, Application US/08738975
; Patent No. 5880267
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Trophinin and trophinin-Assisting
; TITLE OF INVENTION: Proteins
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/738,975
; FILING DATE: Herewith
; CLASSIFICATION: 435

```

```

;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/439,818
; FILING DATE: 05-Dec-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 2251
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2223 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 199..2223
;
; US-08-738-975-8
;
;
; Query Match 52.0%; Score 13; DB 2; Length 2223;
; Best Local Similarity 100.0%; Pred. NO. 1e+02;
; Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 gccgcagcaggtgc 13
; |
; Db 2118 GCCGACGACGTGC 2106
;
;
; RESULT 44
; US-08-728-626-8/C
; Sequence 8, Application US/08728626
; Patent No. 5910451
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting
; TITLE OF INVENTION: Proteins
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,626
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/439,818
; FILING DATE: 12-MAY-1995
; APPLICATION NUMBER: US 08/317,522
; FILING DATE: 04-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 1563
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2223 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

; TOPOLOGY: linear
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 199..2223
 US-08-728-626-8

Query Match 52.0%; Score 13; DB 2; Length 2223;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccgacagagtlgc 13
 |||
 DB 2118 GCGCAGCAGGTGC 2106

RESULT 45
 US-08-808-599A-8/c
 ; Sequence 8, Application US/08808599A
 ; Patent No. 6111089
 ; GENERAL INFORMATION:
 ; APPLICANT: Fukuda, Michiko N.
 ; TITLE OF INVENTION: Trophinin, Trophinin-Assisting
 ; TITLE OF INVENTION: Proteins and Methods to Inhibit Implantation
 ; NUMBER OF SEQUENCES: 41
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Campbell & Flores LLP
 ; STREET: 4370 La Jolla Village Drive, Suite 700
 ; CITY: San Diego
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 92122
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/808,599A
 ; FILING DATE: 28-FEB-1997
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/317,522
 ; FILING DATE: 04-OCT-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/439,818
 ; FILING DATE: 12-MAY-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Campbell, Cathryn A.
 ; REGISTRATION NUMBER: 31,815
 ; REFERENCE/DOCKET NUMBER: P-LA 2256
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 535-9001
 ; TELEFAX: (619) 535-8949
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2223 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 199..2223
 US-08-808-599A-8

Query Match 52.0%; Score 13; DB 3; Length 2223;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccgacagagtlgc 13
 |||
 DB 2118 GCGCAGCAGGTGC 2106

Search completed: December 26, 2001, 12:53:23
 Job time: 6835 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:10:10 : Search time 4619.78 Seconds
(Without alignments)
58.151 Million cell updates/sec

Title: US-09-396-196f-5

Sequence: 1 gcgcagcagcagtcgacgcagcagcatt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 9

Total number of hits satisfying chosen parameters: 600693

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estlom:*
5: em_estlpl:*
6: em_estba:*
7: em_estrov:*
8: em_estrov:*
9: em_hlc:*
10: qb_estl1:*
11: qb_estl2:*
12: qb_hlc:*
13: qb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_liv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	72.0	268	13	BH018624 L18351.d_
2	18	72.0	352	13	BH018623 L1835k.d_
3	18	72.0	370	13	BH018626 L1835m.d_
4	18	72.0	417	13	BH019235 L2544c.d_
5	18	72.0	425	13	BH020126 L6590c.d_
6	18	72.0	456	13	AQ852799 LMAJFV1.1
7	16	64.0	376	11	AQ852799 LMAJFV1.1
8	16	64.0	479	13	D24778 R1CR2545A.R
9	16	64.0	479	13	AQ852582 LMAJFV1.1
10	16	64.0	548	13	AQ852582 LMAJFV1.1
11	16	64.0	690	13	AZ934025 B1_Ba000
12	16	64.0	761	11	Bf864751 963054E05

13	16	64.0	915	11	Bf581354
14	15	60.0	154	11	Bf836559
15	15	60.0	226	13	AQ851231 LMAJFV1.1
16	15	60.0	266	10	A1410504 EST238797
17	15	60.0	352	10	AA680915
18	15	60.0	380	10	AA800788
19	15	60.0	396	10	BE429060
20	15	60.0	418	10	AA625125
21	15	60.0	451	10	AU166733
22	15	60.0	461	13	BH018794
23	15	60.0	463	10	BE466670
24	15	60.0	475	10	AA821080
25	15	60.0	475	11	BG101817
26	15	60.0	487	10	AA698088
27	15	60.0	499	11	Bf916376
28	15	60.0	505	11	C98143
29	15	60.0	514	11	BG467085
30	15	60.0	524	10	A1957283
31	15	60.0	536	11	BG232022
32	15	60.0	545	13	AZ071920
33	15	60.0	553	13	AZ283519
34	15	60.0	566	11	Bf489544
35	15	60.0	567	11	C98142
36	15	60.0	586	13	AZ945591
37	15	60.0	605	13	AQ783596
38	15	60.0	623	11	BG103212
39	15	60.0	730	11	BG479445
40	15	60.0	761	10	BE194330
41	15	60.0	767	11	BE493796
42	15	60.0	792	11	BG299490
43	15	60.0	824	11	BG309144
44	15	60.0	843	10	AJ282984
45	15	60.0	886	10	BE731371
46	15	60.0	893	13	CNSOLJLR
47	15	60.0	898	13	Bf684687
48	15	60.0	979	13	CNSO6ZHL
49	15	60.0	1001	11	BG827666
50	15	60.0	116	10	BE709927
51	14	56.0	119	11	Bf876173
52	14	56.0	178	13	AZ595033
53	14	56.0	196	11	Bf017303
54	14	56.0	204	11	B1229718
55	14	56.0	206	10	A1714341
56	14	56.0	219	10	BB563629
57	14	56.0	226	10	AA326327
58	14	56.0	228	13	TA8C040
59	14	56.0	231	10	A1329089
60	14	56.0	232	11	Bf719645
61	14	56.0	233	11	W83482
62	14	56.0	259	10	AA973902
63	14	56.0	259	10	BE424103
64	14	56.0	267	10	AA515233
65	14	56.0	269	10	BR499261
66	14	56.0	270	10	AA344513
67	14	56.0	275	10	AA261162
68	14	56.0	276	11	F14793
69	14	56.0	279	11	Bf765482
70	14	56.0	286	10	AA165335
71	14	56.0	286	11	W49454
72	14	56.0	287	10	AA260563
73	14	56.0	287	11	Bf765486
74	14	56.0	291	11	BE826582
75	14	56.0	292	10	BB153277
76	14	56.0	293	10	AV119468
77	14	56.0	304	10	A1072655
78	14	56.0	305	10	A1353923
79	14	56.0	305	11	Bf765484
80	14	56.0	306	11	Bf420301
81	14	56.0	309	13	AZ471163
82	14	56.0	309	10	BE595723
83	14	56.0	314	11	B1320345
84	14	56.0	316	10	AA698194
85	14	56.0			

BF581354	602100740
BF836559	CM2-HT096
AQ851231	LMAJFV1.1
A1410504	EST238797
AA680915	LMFAM050
AA800788	EST190285
BE429060	MT0013.H0
AA625125	AT70A04.T
AU166733	AU166733
BH018794	L1975k.d_
BE466670	WHE1139.F
AA821080	GM09622.5
BG101817	RH122.22
AA698088	HL0757.5
Bf916376	IL3-UT011
C98143	C98143.Rice
BG467085	IA04C06.B
A1957283	u185e12.x
BG232022	na135e06.
AZ071920	RPCT-23-3
AZ283519	RPCT-23-1
BF489544	AT25594.5
C98142	C98142.Rice
AZ945591	2M0207H04
AQ783596	HS.3121.A
BG103212	RH122.19
BG479445	60255920
BE194330	HVSMEM008
BF493796	AT02150.5
BG299490	HVSMEM002
BG309144	HVSMEM000
AJ282984	4A3A-P265
BE731371	601565029
AL147200	Anopheles
BF684687	602141056
AL422335	T3 end of
BG827666	602751611
BE709927	IL3-HT061
BF876173	CM4-ET009
AZ595033	1M0407E22
Bf017303	uW7308.X
B1229718	RE27555.5
A1714341	UT-R-AF1-
BB563629	BB563629
AA326327	18754.MAR
AL451592	T.brucel
A1329089	ag12ne.r
R17351	Y913e07.r1
BF719645	maD42g01.
W83482	m128h11.r1
AA973902	oc47g09.s
BE424103	WHE0076.H
AA515233	n669b06.s
BR499261	BB499261
AA344513	24353.MAR
AA261162	0060-3.Lc
F14793	SSC20A05.Po
BF765482	IL2-CS004
AA165335	zq49c06.r
W49454	zEST01191.M
BF765486	IL2-CS004
AA260563	um83c06.Y
BE826582	QV1-EN004
BB153277	BB153277
AV119468	AV119468
A1072655	UT-R-C2-n
A1353923	zebl167.s
Bf765484	IL2-CS004
BF420301	UT-R-B02-
AZ471163	1M0285C12
BE595723	P11-54-10
B1320345	sa121h02.
AA698194	HL03921.5

```

C 86 14 56.0 320 10 AM252743
C 87 14 56.0 322 11 BF714260
C 88 14 56.0 324 11 BE714260
C 89 14 56.0 324 11 BE714022
C 90 14 56.0 324 11 W53911
C 91 14 56.0 328 10 AM478704
C 92 14 56.0 334 13 AA775899
C 93 14 56.0 334 13 BH019034
C 94 14 56.0 335 10 BE655271
C 95 14 56.0 338 13 BH067104
C 96 14 56.0 340 10 AM253547
C 97 14 56.0 341 13 AM902605
C 98 14 56.0 345 11 A2852131
C 99 14 56.0 350 10 BF760406
C 100 14 56.0 352 10 AA326116
      14 56.0 352 10 AF612263

```

ALIGNMENTS

```

RESULT 1
LOCUS BH018624/c
DEFINITION BH018624 268 bp DNA
ACCESSION BH018624
VERSION BH018624.1 GI:14197330
KEYWORDS
SOURCE
ORGANISM Leishmania major.
            Leishmania major
            Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.

```

```

REFERENCE
AUTHORS Myler,P.J., Vogt,C., Cawthra,J., Klacking,M., Marty,A., Mack,J.,
Munden,H., Nguyen,D., Robertson,L., Sisk,E., Fazelinia,G., Aggarwal
,G., Nelson,S., Seyler,A., Worthey,E. and Stuart,K.
TITLE Leishmania major Friedlin Cosmid Genomic Library
JOURNAL Unpublished (2000)
COMMENT Other GSS: L18351.d.HyGT3.1
Contact: Myler PJ
          Seattle Biomedical Research Institute
          4 Nickerson Street, Seattle, WA 98109-1651, USA
          Tel: 206 284-8846
          Fax: 206 284-0313
          Email: mylerp@sbri.org
          Seq primer: HyGT7a
          Class: cosmid ends.

```

```

FEATURES
SOURCE
Location/Qualifiers
1..268
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/clone="L18351"
/clone_lib="Leishmania major Friedlin Cosmid Genomic
Library"
/lab_host="E. coli ED8767"

```

```

/note="Vector: cLHYG; Site:1 BamHI; Genomic DNA from
Leishmania major Friedlin was partially digested with
Sau3AI, size selected, and ligated with BamHI-digested
cLHYG cosmid vector DNA. 9216 clones were picked and
arrayed. Library construction is described in Ivens et
al., Genomics Research, 8:135-145 (1998). The cLHYG
vector (Acc. No. CV059231) is described in Ryan et al.,
Gene, 131:145-150 (1993)."
BASE COUNT 50 a 87 c 91 g 40 t
ORIGIN

```

```

Query Match 72.0%; Score 18; DB 13; Length 268;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ggcgcagcagtgcatcgc 18

```

```

Db 71 GCGCAGCAGTGTCATCGC 54
|||||

```

```

RESULT 2
LOCUS BH018623/c
DEFINITION BH018623 352 bp DNA
ACCESSION BH018623
VERSION BH018623.1 GI:14197329
KEYWORDS
SOURCE
ORGANISM Leishmania major.
            Leishmania major
            Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.

```

```

REFERENCE
AUTHORS Myler,P.J., Vogt,C., Cawthra,J., Klacking,M., Marty,A., Mack,J.,
Munden,H., Nguyen,D., Robertson,L., Sisk,E., Fazelinia,G., Aggarwal
,G., Nelson,S., Seyler,A., Worthey,E. and Stuart,K.
TITLE Leishmania major Friedlin Cosmid End Sequences
JOURNAL Unpublished (2000)
COMMENT Contact: Myler PJ
          Seattle Biomedical Research Institute
          4 Nickerson Street, Seattle, WA 98109-1651, USA
          Tel: 206 284-8846
          Fax: 206 284-0313
          Email: mylerp@sbri.org
          Seq primer: HyGT7a
          Class: cosmid ends.

```

```

FEATURES
SOURCE
Location/Qualifiers
1..352
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/clone="L1835k"
/clone_lib="Leishmania major Friedlin Cosmid Genomic
Library"
/lab_host="E. coli ED8767"

```

```

/note="Vector: cLHYG; Site:1 BamHI; Genomic DNA from
Leishmania major Friedlin was partially digested with
Sau3AI, size selected, and ligated with BamHI-digested
cLHYG cosmid vector DNA. 9216 clones were picked and
arrayed. Library construction is described in Ivens et
al., Genomics Research, 8:135-145 (1998). The cLHYG
vector (Acc. No. CV059231) is described in Ryan et al.,
Gene, 131:145-150 (1993)."
BASE COUNT 63 a 116 c 115 g 55 t 3 others
ORIGIN

```

```

Query Match 72.0%; Score 18; DB 13; Length 352;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ggcgcagcagtgcatcgc 18

```

```

Db 124 GCGCAGCAGTGTCATCGC 107
|||||

```

```

RESULT 3
LOCUS BH018626/c
DEFINITION BH018626 370 bp DNA
ACCESSION BH018626
VERSION BH018626.1 GI:14197332
KEYWORDS
SOURCE
ORGANISM Leishmania major.
            Leishmania major
            Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.

```

```

REFERENCE
1 (bases 1 to 370)

```

AUTHORS
 Myler, P.J., Vogt, C., Cawthra, J., Klacking, M., Marty, A., Mack, J.,
 Munden, H., Nguyen, D., Robertson, L., Sisk, E., Fazelinia, G., Aggarwal,
 G., Nelson, S., Seyler, A., Worthey, E. and Stuart, K.
TITLE
 Leishmania major Friedlin Cosmid End Sequences
JOURNAL
 Unpublished (2000)
COMMENT
 Other_GSSs: L1835m.d_HyGT3.1
 Contact: Myler PJ
 Seattle Biomedical Research Institute
 4 Nickerson Street, Seattle, WA 98109-1651, USA
 Tel: 206 284-8846
 Fax: 206 284-0313
 Email: mylerpj@bri.org
 Seq primer: HyGT7a
 Class: cosmid ends

FEATURES
 source
 Location/Qualifiers
 1..370

/organism="Leishmania major"
 /strain="Friedlin"
 /db_xref="taxon:5664"
 /clone="L1835m"
 /clone_lib="Leishmania major Friedlin Cosmid Genomic
 Library"
 /lab_host="E. coli ED8767"
 /note="Vector: cLHYG; Site_1: BamHI; Genomic DNA from
 Leishmania major Friedlin was partially digested with
 Sau3AI, size selected, and ligated with BamHI-digested
 cLHYG cosmid vector DNA. 9216 clones were picked and
 arrayed. Library construction is described in Ivens et
 al., Genomics Research, 8:135-145 (1998). The cLHYG
 vector (Acc. No. CV059231) is described in Ryan et al.,
 Gene, 131:145-150 (1993)"
 Gene, 131:145-150 (1993)"
 69 a 116 c 128 g 54 t 3 others

BASE COUNT
 Query Match 72.0%; Score 18; DB 13; Length 370;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcgcagcagtgcatgcgc 18
 |||
 Db 66 GCGCAGCAGTGCAATCGC 49

RESULT 4
 BH019235 417 bp DNA GSS 25-MAY-2001
LOCUS
 L2544c.d_HyGT7a.1 Leishmania major Friedlin Cosmid Genomic Library
DEFINITION
 Leishmania major genomic clone L2544c, DNA sequence.
 BH019235
 BH019235.1 GI:14197941
ACCESSION
 GSS.
VERSION
 Leishmania major.
KEYWORDS
 Leishmania major.
SOURCE
 Leishmania major.
ORGANISM
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania.

REFERENCE
 1 (bases 1 to 417)
 Myler, P.J., Vogt, C., Cawthra, J., Klacking, M., Marty, A., Mack, J.,
 Munden, H., Nguyen, D., Robertson, L., Sisk, E., Fazelinia, G., Aggarwal,
 G., Nelson, S., Seyler, A., Worthey, E. and Stuart, K.
 Leishmania major Friedlin Cosmid End Sequences
 Unpublished (2000)
 Other_GSSs: L2544c.d_HyGT3.1
 Contact: Myler PJ
 Seattle Biomedical Research Institute
 4 Nickerson Street, Seattle, WA 98109-1651, USA
 Tel: 206 284-8846
 Fax: 206 284-0313
 Email: mylerpj@bri.org
 Seq primer: HyGT7a
 Class: cosmid ends.

FEATURES
 source
 Location/Qualifiers
 1..417

/organism="Leishmania major"
 /strain="Friedlin"
 /db_xref="taxon:5664"
 /clone="L2544c"
 /clone_lib="Leishmania major Friedlin Cosmid Genomic
 Library"
 /lab_host="E. coli ED8767"
 /note="Vector: cLHYG; Site_1: BamHI; Genomic DNA from
 Leishmania major Friedlin was partially digested with
 Sau3AI, size selected, and ligated with BamHI-digested
 cLHYG cosmid vector DNA. 9216 clones were picked and
 arrayed. Library construction is described in Ivens et
 al., Genomics Research, 8:135-145 (1998). The cLHYG
 vector (Acc. No. CV059231) is described in Ryan et al.,
 Gene, 131:145-150 (1993)"
 Gene, 131:145-150 (1993)"
 78 a 128 c 141 g 66 t 4 others

BASE COUNT
 Query Match 72.0%; Score 18; DB 13; Length 417;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcgcagcagtgcatgcgc 18
 |||
 Db 120 GCGCAGCAGTGCAATCGC 103

RESULT 5
 BH020126/c 425 bp DNA GSS 25-MAY-2001
LOCUS
 L6590c.d_HyGT7a.2 Leishmania major Friedlin Cosmid Genomic Library
DEFINITION
 Leishmania major genomic clone L6590c, DNA sequence.
 BH020126
 BH020126.1 GI:14200134
ACCESSION
 GSS.
VERSION
 Leishmania major.
KEYWORDS
 Leishmania major.
SOURCE
 Leishmania major.
ORGANISM
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania.

REFERENCE
 1 (bases 1 to 425)
 Myler, P.J., Vogt, C., Cawthra, J., Klacking, M., Marty, A., Mack, J.,
 Munden, H., Nguyen, D., Robertson, L., Sisk, E., Fazelinia, G., Aggarwal,
 G., Nelson, S., Seyler, A., Worthey, E. and Stuart, K.
 Leishmania major Friedlin Cosmid End Sequences
 Unpublished (2000)
 Other_GSSs: L6590c.d_HyGT7a.1
 Contact: Myler PJ
 Seattle Biomedical Research Institute
 4 Nickerson Street, Seattle, WA 98109-1651, USA
 Tel: 206 284-8846
 Fax: 206 284-0313
 Email: mylerpj@bri.org
 Seq primer: HyGT7a
 Class: cosmid ends.

FEATURES
 source
 Location/Qualifiers
 1..425

/organism="Leishmania major"
 /strain="Friedlin"
 /db_xref="taxon:5664"
 /clone="L6590c"
 /clone_lib="Leishmania major Friedlin Cosmid Genomic
 Library"
 /lab_host="E. coli ED8767"
 /note="Vector: cLHYG; Site_1: BamHI; Genomic DNA from
 Leishmania major Friedlin was partially digested with
 Sau3AI, size selected, and ligated with BamHI-digested
 cLHYG cosmid vector DNA. 9216 clones were picked and
 arrayed. Library construction is described in Ivens et
 al., Genomics Research, 8:135-145 (1998). The cLHYG
 vector (Acc. No. CV059231) is described in Ryan et al.,
 Gene, 131:145-150 (1993)"
 Gene, 131:145-150 (1993)"

BASE COUNT
 89 a 129 c 137 g 70 t

ORIGIN

Query Match
Best Local Similarity 72.0%; Score 18; DB 13; Length 425;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcagcaggtgcatcgc 18
|||||
DB 131 GCGCAGCAGGTGTCATCGC 114

RESULT 6

LOCUS A0852799 456 bp DNA GSS 25-MAY-2001
DEFINITION LMAJFV1.Lm77909.y1 Leishmania major FV1 random genomic library

ACCESSION LMAJFV1.Lm77909.5 similar to
VERSION TR:062388 062388 ATAXIA TELANGICTASIA PROTEIN MUTATED IN HUMAN
KEYWORDS BEINGS ; DNA sequence.

A0852799
A0852799.1 GI:6119124
GSS.

SOURCE Leishmania major.
ORGANISM Leishmania major
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 456)
AUTHORS Akopyants, N.S., Clifton, S.W., Martin, J., Pape, D., Wylie, T., Li, L.,

Kissinger, J., Roos, D.S., Marra, M., Hillier, L., Chinwalla, A.,
Blistein, A., Schmitt, A., Person, B., Theising, B., Ritter, E., Ronko,
J., Bennett, J., Cole, R., Underwood, K., Cardenas, M., Gibbons, M.,
Harvey, N., McCann, R., Tsagaris, R., Williams, T., Jackson, V.,
Bowers, Y., Swaller, T., Waterston, R., Wilson, R., and Beverley, S.M.
A survey of the Leishmania major Friedlin Strain VI genome by
shotgun sequencing: a resource for DNA microarrays and expression
profiling
Mol Biochem. Parasitol. 113 (2), 337-340 (2001)

JOURNAL MEDLINE 21192569
COMMENT Contact: Akopyants, NS / Beverley, SM
WashU Leishmania Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewartson.wustl.edu

Library construction: Natalia S. Akopyants, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
If using this information please cite:
N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
Friedlin strain VI genome by shotgun sequencing' and the Washington
University Genome Sequencing Center for information on obtaining
clone material please contact: Natalia S. Akopyants Ph.D.
(natalia@porcim.wustl.edu) and/or Stephen M. Beverley Ph.D.
(beverley@porcim.wustl.edu)

Seq primer: -40RP from GIBCO
Class: shotgun
High quality sequence stop: 252.
Location/Qualifiers
1..456

FEATURES

Source

/organism="Leishmania major"
/strain="Friedlin strain VI"
/db_xref="taxon:5664"
/clone="LMAJFV1.Lm77909"
/clone_lib="Leishmania major FV1 random genomic library"
/lab_hosts="TOP10 (Invitrogen)"
/note="Vector: pZero-2 (Invitrogen); Site 1: EcoRV;
Genomic DNA was isolated from stationary phase cells. For
this library, DNA was sheared to give a tight size
distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
polymerase, dephosphorylated with Shrimp Alkaline
Phosphatase and ligated into pZero-2 vector's EcoRV site."
78 a 141 c 144 g 93 t

BASE COUNT
ORIGIN

Query Match
Best Local Similarity 72.0%; Score 18; DB 13; Length 456;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcagcaggtgcatcgc 18
|||||
DB 253 GCGCAGCAGGTGTCATCGC 270

RESULT 7

LOCUS BF933544 376 bp mRNA EST 22-JAN-2001
DEFINITION CM1-NT0270-211200-661-903 NT0270 Homo sapiens CDNA, mRNA sequence.

ACCESSION BF933544
VERSION BF933544.1 GI:12350868
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 376)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663

JOURNAL MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?pl=CM1&t2=CM1-NT0270-
211200-661-903&t3=2000-12-21&t4=1)

Seq primer: puc 18 forward
High quality sequence start: 10
High quality sequence stop: 376.
Location/Qualifiers
1..376

FEATURES

Source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NT0270"
/dev_stage="Adult"
/note="Organ: nervous_tumor; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. letters patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
low stringency conditions."
77 a 104 c 99 g 96 t

BASE COUNT
ORIGIN

Query Match
Best Local Similarity 64.0%; Score 16; DB 11; Length 376;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ggtgcatcgcagcat 24
|||||
DB 329 GGTGCAATCGCAGCAT 314

BASE COUNT
ORIGIN

RESULT 8
 D24778 399 bp mRNA EST 08-JUL-1999
 LOCUS R1C2545A Rice root Oryza sativa cDNA clone R2545_1A, mRNA
 DEFINITION sequence.
 ACCESSION D24778 GI:428626
 VERSION D24778.1 GI:428626
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa.
 Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 399)
 Minobe, Y. and Sasaki, T.
 Rice cDNA from root
 Unpublished (1995)
 CONTACT: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
 305-8602, Japan
 Tel: 81-298-38-7441
 Fax: 81-298-38-7468
 Email: tsasaki@agr.affrc.go.jp, URL: http://rnp.dna.affrc.go.jp/
 PROJECT = 'RGP'.

FEATURES
 Source location/Qualifiers
 1..399
 /organism="Oryza sativa"
 /strain="Nipponbare, sub-species Japonica"
 /db_xref="taxon:4530"
 /clone_lib="R2545_1A"
 /clone_1ib="Rice root"
 /note="Prepared from seedling root."

BASE COUNT 114 a 70 c 102 g 112 t 1 others

ORIGIN

Query Match 64.0% Score 16; DB 11; Length 399;
 Best Local Similarity 100.0% Pred. No. 1.3e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 ggtgcatcgccagcat 24
 ||||||||||||||||
 Db 251 ggtgcatcgccagcat 266

RESULT 9
 A0852582 479 bp DNA GSS 25-MAY-2001
 LOCUS LMAJFV1_lm74c04.y1 Leishmania major FV1 random genomic library
 DEFINITION Leishmania major genome clone LMAJFV1_lm74c04 5', DNA sequence.
 ACCESSION A0852582
 VERSION A0852582.1 GI:6118907
 KEYWORDS GSS.
 SOURCE Leishmania major.
 ORGANISM Leishmania major.
 Eukaryota: Euklenozoa: Kinetoplastida; Trypanosomatidae;
 Leishmania.
 1 (bases 1 to 479)
 Akopyants, N.S., Clifton, S.W., Martin, J., Pape, D., Wylie, T., Li, L.,
 Kistinger, J., Roos, D.S., Marra, M., Hillier, L., Chinwalla, A.,
 Blistan, A., Schmitt, A., Persen, B., Theising, B., Rifter, E., Ronko,
 I., Bennett, J., Cole, R., Underwood, K., Cardenas, M., Gibbons, M.,
 Harvey, N., McCann, R., Tsagarelis, H., Williams, T., Jackson, Y.,
 Bowers, Y., Swaller, T., Waterston, R., Wilson, R. and Beverley, S. M.
 A survey of the Leishmania major Friedlin strain VI genome by
 shotgun sequencing: a resource for DNA microarrays and expression
 profiling
 Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)
 21192569
 Contact: Akopyants, NS / Beverley, SM
 Washu Leishmania Project

JOURNAL
 MEDLINE
 COMMENT

Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 Library construction: Natalia S. Akopyants, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 It using this information please cite:
 N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
 Friedlin strain VI genome by shotgun sequencing' and the Washington
 University Genome Sequencing Center for information on obtaining
 clone material please contact: Natalia S. Akopyants Ph.D.
 (natalia@wustl.wustl.edu) and/or Stephen M. Beverley Ph.D.
 (beverley@wustl.wustl.edu)
 Seq primer: -40RP from Gibco
 Class: shotgun
 High quality sequence stop: 372.

FEATURES
 Source location/Qualifiers
 1..479
 /organism="Leishmania major"
 /strain="Friedlin strain VI"
 /db_xref="taxon:5664"
 /clone="LMAJFV1_lm74c04"
 /clone_1ib="Leishmania major FV1 random genomic library"
 /lab_host="TOP10 (Invitrogen)"
 /note="Vector: pZero-2 (Invitrogen); Site: 1: EcoRV;
 Genomic DNA was isolated from stationary phase cells. For
 this library, DNA was sheared to give a tight size
 distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
 polymerase, dephosphorylated with Shrimp Alkaline
 Phosphatase and ligated into pZero-2 vector's EcoRV site."

BASE COUNT 98 a 155 c 147 g 79 t

ORIGIN

Query Match 64.0% Score 16; DB 13; Length 479;
 Best Local Similarity 100.0% Pred. No. 1.3e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 cagcagtcgcatgcc 19
 ||||||||||||||||
 Db 47 cagcagtcgcatgcc 62

RESULT 10
 P968R 548 bp DNA GSS 25-JUL-2000
 LOCUS Leishmania major Friedlin PAC P968 right end-sequence, genomic
 DEFINITION survey sequence.
 ACCESSION AL390674
 VERSION AL390674.1 GI:9501650
 KEYWORDS GSS.
 SOURCE Leishmania major.
 ORGANISM Leishmania major.
 Eukaryota: Euklenozoa: Kinetoplastida; Trypanosomatidae;
 Leishmania.
 1 (bases 1 to 548)
 Ivens, A.C., Lewis, S.M., Bagherzadeh, A., Zhang, L., Chan, H.M. and
 Smith, D.F.
 A physical map of the Leishmania major Friedlin genome
 Genome Res. 8 (2), 135-145 (1998)
 98146435
 2 (bases 1 to 548)
 Taylor, R.G., Buckle, E.E.J., Ivens, A.C., Rajandream, M.A. and
 Barrell, B.G.
 Direct Submission
 Submitted (24-JUL-2000) Leishmania major Friedlin genome sequencing
 project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
 Cambridge CB10 1SA, E-mail: barrel@sanger.ac.uk and
 alicatesanger.ac.uk
 see http://www.edi.ac.uk/parasites/leish.html
 Details of Leishmania sequencing at the Sanger Centre are available
 at http://www.sanger.ac.uk/projects/L_majior/

JOURNAL
 MEDLINE
 COMMENT

The primer sequence can be obtained from alicat@sanger.ac.uk.

FEATURES
Location/Qualifiers
source
1..548
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/clone="PAC P968"

BASE COUNT
86 a 167 c 163 g 132 t

Query Match
Best Local Similarity 100.0%; Score 16; DB 13; Length 548;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 gtcacgcgcacgacat 25
|||||

Db 541 GTGCATCGCCACATT 526

RESULT 11
A2934025/c

DEFINITION
BJ_B00001N11f B. japonicum BAC library Bradyrhizobium japonicum
genomic, DNA sequence.

ACCESSION
A2934025
VERSION
A2934025.1 GI:13776085

KEYWORDS
GSS.

SOURCE
Bradyrhizobium japonicum.

ORGANISM
Bradyrhizobium japonicum.

REFERENCE
1 (bases 1 to 690)
Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A., Goicoechea
A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
genome

TITLE
Genome Res. 11 (8), 1434-1440 (2001)

JOURNAL
MEDIINE
21376150

COMMENT

Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Class: BAC ends
High quality sequence start: 3
High quality sequence stop: 650.

FEATURES

Location/Qualifiers

1..690

/organism="Bradyrhizobium japonicum"

/strain="JUSDA110"

/db_xref="taxon:375"

/clone_11b="B. japonicum BAC library"

/lab_host="E. coli"

/note="Vector: pIndigo536; Site_1: HindIII"

BASE COUNT
135 a 200 c 220 g 134 t 1 others

ORIGIN

Query Match
Best Local Similarity 100.0%; Score 16; DB 13; Length 690;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 gacagcagtgatcgc 18
|||||

Db 621 GCAGCAGTGATCCG 606

RESULT 12
BF864751/c

LOCUS

BF864751 761 bp mRNA EST

19-JAN-2001

DEFINITION

963054E05.Y1 C. reinhardtii CC-1690, Stress condition I, normalized
lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION
BF864751
VERSION
BF864751.1 GI:12254895

KEYWORDS

SOURCE

ORGANISM

Chlamydomonas reinhardtii.
Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.

REFERENCE

AUTHORS

TITLE

Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants: project phase 3

Unpublished (2000)
Contact: Charles Hauser
DCMB Box 91000

Duke University
Durham, NC 27708-1000

Tel: 919 613 8159
Fax: 919 613 8177

Email: chauser@duke.edu.

JOURNAL

COMMENT

source

location/Qualifiers

1..761

/organism="Chlamydomonas reinhardtii"

/strain="CC-1690 wild type mt+ 21gr"

/db_xref="taxon:3055"

/clone_11b="C. reinhardtii CC-1690, Stress condition I,
normalized, lambda zap II"

/note="Vector: pBluescript II SK-; site_1: EcoRI; site_2:
XhoI; This library, constructed by John Davies and Jeffrey

McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr

, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA RNA was
purified from each sample, pooled and cDNA synthesized.

The cDNA was directionally cloned into lambda zap II
(Stratagene) in the EcoRI (5') and XhoI (3') sites.

Plasmids were excised from the lambda
zap clones by superinfection with Exsist (Stratagene)

phage. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT
153 a 179 c 308 g 117 t 4 others

ORIGIN

Query Match
Best Local Similarity 100.0%; Score 16; DB 11; Length 761;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 aggtcgcgcgcacga 23
|||||

Db 466 AGGTCGATCGCCACGA 451

RESULT 13
BF581354

LOCUS

BF581354

DEFINITION

602100740F1 NC1_CGAP_Co24 Mus musculus cDNA clone IMAGE:422425 5',

mRNA sequence.

ACCESSION
BF581354

VERSION
BF581354.1 GI:11655066

KEYWORDS

SOURCE

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 915)
NIH-MGC http://mgi.nhl.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.

/db_xref="taxon:5664"
 /clone_lib="Leishmania major FV1 random genomic library"
 /lab_host="TOP10 (Invitrogen)"
 /note="Vector: pZero-2 (Invitrogen); Site_1: EcoRI;
 Genomic DNA was isolated from stationary phase cells. For
 this library, DNA was sheared to give a tight size
 distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
 polymerase, dephosphorylated with Shrimp Alkaline
 Phosphatase and ligated into pZero-2 vector's EcoRV site."

BASE COUNT

40 a 67 c 64 g 55 t

ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 226;
 Best Local Similarity 100.0%; Pred. No. 4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 88 CGCAGCAGTGCATC 102

RESULT 16
 LOCUS A1410504 266 bp mRNA EST 09-FEB-1999
 DEFINITION ESR238797 Normalized rat heart, Bento Soares Rattus sp. cDNA clone
 ACCESSION RHDCX29 3' end, mRNA sequence.
 VERSION A1410504
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 266)
 Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
 Kerlavage, A.R. and Adams, M.D.
 Rat Genome Project: Generation of a Rat EST (RESt) Catalog & Rat
 Gene Index
 Unpublished (1998)
 CONTACT: Lee, NH
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13-21.

FEATURES

Source Location/Qualifiers
 1..266
 /organism="Rattus sp."
 /db_xref="taxon:10118"
 /clone="RHDCX29"
 /clone_lib="Normalized rat heart, Bento Soares"
 /note="Organ: heart; Vector: pTY73Pac; Site_1: EcoRI;
 Site_2: NotI"
 BASE COUNT 56 a 61 c 64 g 85 t
 ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 266;
 Best Local Similarity 100.0%; Pred. No. 4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 90 GGTGATGCGCAGCA 104

RESULT 17
 LOCUS AA680915/c 352 bp mRNA EST 10-DEC-1998

DEFINITION LmfAm0507 Leishmania major Amastigote full length cDNA library
 ACCESSION AA680915
 VERSION AA680915.1 GI:3886275
 KEYWORDS EST.
 SOURCE Leishmania major.
 ORGANISM Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania.
 1 (bases 1 to 352)
 Norrish, A.R., Dyall, S.D., Smith, D.F. and Blackwell, J.M.
 Analysis of Leishmania Major Amastigote Expressed Sequence Tags
 Unpublished (1997)
 On Dec 5, 1997 this sequence version replaced gi:2662920.
 CONTACT: Blackwell, JM
 Cambridge Institute for Medical Research
 Wellcome Trust/MRC Building, Addenbrooke's Hospital, Hills Road,
 Cambridge CB2 2XY, UK
 Tel: 01223 336 143
 Fax: 01223 331 206
 Email: jmb37@cus.cam.ac.uk
 PCR PRIMERS
 FORWARD: GTAAACGACGGCCAGT
 BACKWARD: GGAACAGCTATGACCATG
 Seq primer: AATTACCTCCTAAGAGG
 High quality sequence stop: 352.
 Location/Qualifiers
 1..352
 /organism="Leishmania major"
 /strain="Friedlin"
 /db_xref="taxon:5664"
 /clone="O83"
 /clone_lib="Leishmania major Amastigote full length cDNA
 library"

BASE COUNT 70 a 110 c 86 g 86 t

Query Match 60.0%; Score 15; DB 10; Length 352;
 Best Local Similarity 100.0%; Pred. No. 4.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 183 CAGCAGTGCATCGC 169

RESULT 18
 LOCUS AA800788 380 bp mRNA EST 30-APR-1998
 DEFINITION ESR190285 Normalized rat lung, Bento Soares Rattus sp. cDNA clone
 ACCESSION RUDAL45 3' end, mRNA sequence.
 VERSION AA800788
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 380)
 Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
 Kerlavage, A.R. and Adams, M.D.
 Rat Genome Project: Generation of a Rat EST (RESt) Catalog & Rat
 Gene Index
 Unpublished (1998)
 CONTACT: Lee, NH
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13-21.

Query Match 60.0%; Score 15; DB 10; Length 352;
 Best Local Similarity 100.0%; Pred. No. 4.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 183 CAGCAGTGCATCGC 169

RESULT 18
 LOCUS AA800788 380 bp mRNA EST 30-APR-1998
 DEFINITION ESR190285 Normalized rat lung, Bento Soares Rattus sp. cDNA clone
 ACCESSION RUDAL45 3' end, mRNA sequence.
 VERSION AA800788
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 380)
 Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
 Kerlavage, A.R. and Adams, M.D.
 Rat Genome Project: Generation of a Rat EST (RESt) Catalog & Rat
 Gene Index
 Unpublished (1998)
 CONTACT: Lee, NH
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13-21.


```

FEATURES
  source
    Location/Qualifiers
      1..380
        /organism="Rattus sp."
        /db_xref="taxon:10118"
        /clone_1b="Normalised rat lung, Bento Soares"
        /note="Organ: lung; Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT
  71 a 100 c 106 g 103 t
ORIGIN
  Query Match
    Best Local Similarity 100.0%; Score 15; DB 10; Length 380;
    Pred. No. 4,1e+02; Mismatches 0; Indels 0; Gaps 0;
  Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 ggtgcatcgccagca 23
    |||||
Db 41 GGTGATCGCCAGCA 55

RESULT 19
BE429060 396 bp mRNA EST 26-JUL-2000
LOCUS BE429060.1 H06F990621 ITFC MTD Durum Wheat Root Library Trilicium
DEFINITION turgidum subsp. durum cDNA clone MTD013.H06, mRNA sequence.
ACCESSION BE429060
VERSION BE429060.1 GI:9426903
KEYWORDS durum wheat.
SOURCE Trilicium turgidum subsp. durum
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
: Triliceae; Trilicium.
1 (bases 1 to 396)
Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
Pechiolli,N., Quiset,C., Schuch,W., Selvaraj,G., Shalhoub,M.,
Sorrenti,M., Warburton,M., and Wenzel,G.
International Triliceae EST Cooperative (ITFC): Production of
Expressed Sequence Tags for Species of the Triliceae
Unpublished (2000)
Contact: Joudrier P
INRA, Unite de Biochimie et Biologie Moleculaire des Cereales
2, place VIALA, 34060 Montpellier cedex 01 FRANCE
Tel: 33 4 99 61 23 84
Fax: 33 4 99 61 23 48
Email: joudrier@enscm.inra.fr
International Triliceae EST Cooperative (ITFC)
http://wheat.pw.usda.gov/genome.
FEATURES
  source
    Location/Qualifiers
      1..396
        /organism="Trilicium turgidum subsp. durum"
        /cultivar="Silliana"
        /db_xref="taxon:4567"
        /clone="MTD013.H06"
        /clone_1b="ITFC MTD Durum Wheat Root Library"
        /tissue_type="root"
        /dev_stage="3-day-old seedling, water-stressed"
        /note="Vector: pSPORT1; T7 primers used. See pSPORT1
        polylinker site. 0.3-2.0 kbp average insert size."
BASE COUNT
  96 a 87 c 110 g 97 t
ORIGIN
  Query Match
    Best Local Similarity 100.0%; Score 15; DB 10; Length 396;
    Pred. No. 4,1e+02; Mismatches 0; Indels 0; Gaps 0;
  Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 gcagtgatcgccca 20

```

```

Db 75 GCAGTGATCGCCCA 61
|||||
RESULT 20
AA625125 418 bp mRNA EST 02-MAR-1998
LOCUS AA625125 aF0a04.0 r1 Soares_NhMpu_S1 Homo sapiens cDNA clone IMAGE:1047342
DEFINITION 5', mRNA sequence.
ACCESSION AA625125
VERSION AA625125.1 GI:2537510
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 418)
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
CONTACT: Wilson RK
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 626 Std Error: 0.00
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 408.
FEATURES
  source
    Location/Qualifiers
      1..418
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:1047342"
        /clone_1b="Soares_NhMpu_S1"
        /tissue_type="Pooled human melanocyte, fetal heart, and
        pregnant uterus"
        /lab_host="DH10B"
        /note="Organ: mixed (see below); Vector: pT7T3D-Pac
        (Pharmacia) with a modified polylinker; Site_1: Not I;
        Site_2: Eco RI; Equal amounts of plasmid DNA from three
        normalized 11: uries (melanocyte 2NBH, pregnant uterus
        NbHpu), and fetal heart NbH19W) were mixed, and ss circles
        were made in vitro. Following NHP purification, this DNA
        was used as tracer in a subtractive hybridization
        reaction. The driver was PCR-amplified cDNAs from pools of
        5,000 clones made from the same 3 libraries. The pools
        consisted of T.M.A.G.E. clones 260232-265223,
        340488-345479, and 484488-489473."
BASE COUNT
  128 a 93 c 114 g 83 t
ORIGIN
  Query Match
    Best Local Similarity 100.0%; Score 15; DB 10; Length 418;
    Pred. No. 4,1e+02; Mismatches 0; Indels 0; Gaps 0;
  Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ccagcaggtgcatc 16
    |||||
Db 399 CGAGGAGGTGATC 413

RESULT 21
AU166733 453 bp mRNA EST 26-APR-2001
LOCUS AU166733 Rice callus (2001) Oryza sativa cDNA clone C50066, mRNA
DEFINITION sequence.
ACCESSION AU166733

```

```

VERSION      AU166733.1  GI:12405132
KEYWORDS
SOURCE       Oryza sativa
ORGANISM     Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
              Ehrhartoideae; Oryzoideae; Oryza.
REFERENCE    1 (bases 1 to 453)
AUTHORS      Sasaki, T. and Yamamoto, K.
TITLE        Rice cDNA from callus (2001)
JOURNAL      Unpublished (2001)
COMMENT      Contact: Takuji Sasaki
              National Institute of Agrobiological Resources
              Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
              305-8602, Japan
              Tel: 81-298-38-7441
              Fax: 81-298-38-7468
              Email: tsasaki@abrr.affrc.go.jp, URL: http://rjp.dna.affrc.go.jp/
              PROJECT="RGP".
              C50066_982.

FEATURES
  source      1..453
              /organism="Oryza sativa"
              /cultivar="Nipponbare"
              /db_xref="taxon:4530"
              /clone_lib="Rice callus (2001)"
              /tissue_type="callus"

BASE COUNT   164 a      83 c      140 g      66 t

ORIGIN
Query Match  Best Local Similarity  60.0%; Score 15; DB 10; Length 453;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  11 tgcacgcacacat 25
    |||||||||||
Db  349 TGCAATCGCCACAT 335

RESULT 22
LOCUS      BH018794/c  461 bp      DNA      GSS      25-MAY-2001
DEFINITION L1975K.d.HyGT3.1 Leishmania major Friedlin Cosmid Genomic Library
ACCESSION  BH018794
VERSION    BH018794.1  GI:14197500
KEYWORDS  GSS.
SOURCE     Leishmania major.
           Leishmania major
           Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
           Leishmania.
REFERENCE  1 (bases 1 to 461)
AUTHORS    Myler, P.J., Vogt, C., Cawthra, J., Kieckheaf, M., Marty, A., Mack, J.,
           Munden, H., Nguyen, D., Robertson, L., Sisk, E., Fazellina, G., Aggarwal,
           G., Nelson, S., Seyler, A., Worthey, E. and Stuart, K.
TITLE      Leishmania major Friedlin Cosmid End Sequences
JOURNAL    Unpublished (2000)
COMMENT    Other GSS: L1975K.d.HyGT7A.1
           Contact: Myler PJ
           Seattle Biomedical Research Institute
           4 Nickerson Street, Seattle, WA 98109-1651, USA
           Tel: 206 284-8846
           Fax: 206 284-0313
           Email: mylerpj@sbri.org
           Seq primer: HyGT3
           Class: cosmid ends.

FEATURES
  source      Location/Qualifiers
              1..461
              /organism="Leishmania major"
              /strain="Friedlin"
              /db_xref="taxon:5664"

```

```

BASE COUNT   100 a      130 c      129 g      101 t      1 others

ORIGIN
Query Match  Best Local Similarity  60.0%; Score 15; DB 13; Length 461;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  7 caggtgcacgcacg 21
    |||||||||||
Db  279 CAGGTGCATGCCACG 265

RESULT 23
LOCUS      BE446670/c  463 bp      mRNA      EST      25-JUL-2000
DEFINITION WHE1139.F07.L13Z5 wheat etiolated seedling root normalized cDNA
           library Triticum aestivum cDNA clone WHE1139.F07.L13, mRNA
           sequence.
ACCESSION  BE446670
VERSION    BE446670.1  GI:9446232
KEYWORDS  EST.
SOURCE     Bread wheat.
           Triticum aestivum
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae
           ; Triticeae; Triticum.
REFERENCE  1 (bases 1 to 463)
AUTHORS    Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han,
           P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Nguyen, H.T.,
           Rausch, C.J., Seaton, C.L., Tong, J.C. and Zhang, D.
TITLE      The structure and function of the expressed portion of the wheat
           genomes - Normalized root cDNA library
JOURNAL    Unpublished (2000)
COMMENT    Contact: Olin Anderson
           US Department of Agriculture, Agriculture Research Service, Pacific
           West Area, Western Regional Research Center
           800 Buchanan Street, Albany, CA 94710, USA
           Tel: 5105595773
           Fax: 5105595818
           Email: oanderson@nwp.usda.gov
           Sequence have been trimmed to remove vector sequence and low
           quality sequence with phred score less than 20
           Seq primer: StrataGene SK primer.

FEATURES
  source      Location/Qualifiers
              1..463
              /organism="Triticum aestivum"
              /cultivar="Chinese Spring"
              /db_xref="taxon:4565"
              /clone_lib="WHE1139_F07_L13"
              /clone_lib="Wheat etiolated seedling root normalized cDNA
              library"
              /tissue_type="Root"
              /dev_stage="Five day old etiolated seedling"
              /lab_host="E. coli DH10B"
              /note="Vector: Lambda Uni-ZAP XR, excised phagemid
              pBluescript SK, Site 1; EcoRI: Site 2; XhoI: Seeds were
              surface-sterilized, germinated and grown aseptically in
              the dark at room temperature on filter paper with water,
              mystatin and cefotaxime in covered crystallization
              dishes. Roots were harvested. The tissue, total RNA, and

```

poly(A) RNA were prepared, a cDNA library was made in the T7 Close Lab (Choi, Close, Penton) at the University of California, Riverside. The cDNA clones were in vivo excised to give pluscript phagemids before normalization. The mass excision of normal library and normalization were done in HT Nguyen lab by D. Zhang at Texas Tech University. Normalization protocol used was that of Soares. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT

119 a 98 c 132 g 114 t

Query Match 60.0%; Score 15; DB 10; Length 463;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 gcaagtgcaccca 20
|||||
Db 111 GCAGTGCATCGCCA 97

RESULT 24
AA821080 475 bp mRNA EST 23-APR-2001
LOCUS
DEFINITION GM09622.5prime GM Drosophila melanogaster ovary Bluescript
Drosophila melanogaster cDNA clone GM09622 5prime, mRNA sequence.

ACCESSION AA821080
VERSION AA821080.1 GI:2890948
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 475)
HARVEY, D., BROKSTEIN, P., HONG, L., EVANS-HOLM, M., SU, C., TSANG, G., LEWIS, S. and RUBIN, G.M.
BDGP/HMT Drosophila EST Project
Unpublished (2001)
TITLE JOURNAL
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
Plate: 96 row: B column: 10
High quality sequence stop: 439.

FEATURES
source.
1..475
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="GM09622"
/clone_lib="GM Drosophila melanogaster ovary Bluescript"
/sex="female"
/dev_stage="newly eclosed females: germlarium-stage 6"
/lab_host="SOLR"
/note="Organ: ovary; Vector: Bluescript SK; Site: 1; EcoRI; Site: 2; XhoI; Constructed using Stratagene ZAP-CDNA Synthesis Kit. Oligo dT-primed and directionally cloned at EcoRI and XhoI in Bluescript SK(+/-)"

BASE COUNT 128 a 147 c 121 g 79 t

Query Match 60.0%; Score 15; DB 10; Length 475;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 aggtgcaccca 22
|||||
Db 344 AGTGCATCGCCACG 358

RESULT 25
BG101817 475 bp mRNA EST 30-JAN-2001
LOCUS
DEFINITION RH122.22.G09.g1.A003 Rhizome2 (RH122) Sorghum prolinguam cDNA, mRNA sequence.
BG101817
ACCESSION BG101817.1 GI:12616762
VERSION
KEYWORDS
SOURCE Sorghum prolinguam.
ORGANISM Sorghum prolinguam
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 475)
CORDONIER-PRATT, M.-M., GINGLE, A., PATERSON, A., SUDMAN, M. and PRATT, L.H.
An EST database from Sorghum: Sorghum prolinguam rhizomes Unpublished (2000)
CONTACT: Cordonier-Pratt MM
TITLE JOURNAL
COMMENT Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Seq primer: PolyTmix
High quality sequence start: 11
High quality sequence stop: 408
POLYA-NO.

FEATURES
source
1..475
/organism="Sorghum prolinguam"
/db_xref="taxon:132711"
/clone_lib="Rhizome2 (RH122)"
/note="Organ: Rhizomes; Vector: pluscript II from Lambda Zap II; Site: 1; XhoI; Site: 2; EcoRI; The library was made from poly-A RNA in the cloning vector Lambda Zap II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 128 a 107 c 113 g 127 t

Query Match 60.0%; Score 15; DB 11; Length 475;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 gcaagtgcaccca 20
|||||
Db 240 GCAGTGCATCGCCA 254

RESULT 26
AA698088 487 bp mRNA EST 19-APR-2001
LOCUS
DEFINITION HL03757.5prime HL Drosophila melanogaster head Bluescript
Drosophila melanogaster cDNA clone HL03757 5prime, mRNA sequence.
AA698088
ACCESSION AA698088.1 GI:2701017
VERSION
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 487)
HARVEY, D., BROKSTEIN, P., HONG, L., EVANS-HOLM, M., SU, C., TSANG, G., LEWIS, S. and RUBIN, G.M.
BDGP/HMT Drosophila EST Project
Unpublished (2001)
TITLE JOURNAL
COMMENT Contact: Stapleton, M.
BDGP

Lawrence Berkeley National Lab
One Cyclotron Rd
Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
Plate: 37 row: E column: 9
High quality sequence stop: 331.
Location/Qualifiers

FEATURES

source

```

1..487
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="HL03757"
/clone_lib="HL Drosophila melanogaster head Bluescript"
/sex="male and female"
/dev_stage="adult"
/lab_host="SOLR"
/note="organ: head-brain & sensory organ; Vector:
Bluescript SK; Site_1: EcoRI; Site_2: XhoI; Constructed
using Stratagene Zap-cDNA Synthesis Kit. Oligo dT-primed
and directionally cloned at EcoRI and XhoI in Bluescript
SK(+/-)"
BASE COUNT      123 a      151 c      128 g      85 t
ORIGIN

```

Query Match 60.0%; Score 15; DB 10; Length 487;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 agcaagtcacgcgc 19
|||||
Db 109 AGCAGTCATCGCC 95

RESULT 27
LOCUS BP916376 499 bp mRNA EST 18-JAN-2001
DEFINITION IL3-UT0115-111200-377-F05 UT0115 Homo sapiens cDNA, mRNA sequence.
ACCESSION BP916376
VERSION BP916376.1 GI:12307834
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 499)
Dias Neto E., Garcia Correa R., Verjovski-Almeida S., Briones M.R.,
Nagai M.A., da Silva W. Jr., Zagro M.A., Bordin S., Costa F.F.,
Goldman G.H., Carvalho A.F., Matsukuma A., Baia G.S., Simpson D.H.,
Brunstein A., deoliveira P.S., Bucher P., Jongeneel C.V., O'Hare
M.J., Soares F., Brentani R.R., Reis L.F., de Souza S.J. and
Simpson A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=IL3&f2=IL3-UT0115-111200-377-F05&f3=2000-12-11&f4=1>)
Seq primer: puc 18 forward
High quality sequence stop: 496.
Location/Qualifiers

FEATURES

source

```

1..499
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```

/clone_lib="UT0115"
/dev_stage="Adult"
/note="organ: uterus-tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (0.5. letters Patent application
No. 196/716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      121 a      121 c      114 g      143 t
ORIGIN

```

Query Match 60.0%; Score 15; DB 11; Length 499;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 gccagtcacgcgcca 20
|||||
Db 216 GCAGTCATCGCCA 230

RESULT 28
LOCUS C98143 505 bp mRNA EST 29-APR-1999
DEFINITION C98143 Rice callus Oryza sativa cDNA clone C0777_7A, mRNA sequence.
ACCESSION C98143 D28202
VERSION C98143.1 GI:3760889
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 505)
Yamamoto K. and Sasaki T.
Rice cDNA from callus (1998)
Unpublished (1998)
On Oct 19, 1998 this sequence version replaced gi:454479.
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@db.affrc.go.jp, URL: <http://rpg.dna.affrc.go.jp/>
D28202: Submitted (01-JAN-1994)
PROJECT = "RGP"
POLY = No.

FEATURES

source

```

1..505
/organism="Oryza sativa"
/strain="cultivar Nipponbare, sub-species Japonica"
/db_xref="taxon:4530"
/clone="C0777_7A"
/clone_lib="Rice callus"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid."
BASE COUNT      158 a      103 c      170 g      72 t
ORIGIN

```

Query Match 60.0%; Score 15; DB 11; Length 505;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 tgcacgcacgacatt 25
|||||
Db 475 TGACCGCCACGACATT 461

RESULT	29
LOCUS	BG467085
DEFINITION	BCA67085 514 bp mRNA EST 20-MAR-2001 1A04C06 Bovine Mixed Adipose cDNA library Bos taurus cDNA 5', mRNA sequence.
ACCESSION	BG467085
VERSION	BG467085.1 GI:13396060
KEYWORDS	EST.
SOURCE	cov.
ORGANISM	Bos taurus
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae 1 (bases 1 to 514) Moore,S.S., Hansen,C., Li,C., Fu,A., Meng,Y., Li,G., Murdoch,G., Dixon,W. and Christopherson,B. cDNA's from bovine subcutaneous adipose tissue unpublished (2001)
TITLE	Contact: Dr. Stephen Moore
JOURNAL	. Beef Genomics Laboratory
COMMENT	Dept of AFNS, University of Alberta 410 Agri/For. Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada Tel: 780 492 0169 Fax: 780 492 4265 Email: smoores@atns.ualberta.ca The sequence best matches gb:BTNDHDUB (B.taurus CI-B14.5b mRNA form NADH dehydrogenase (ubiquinone)) in main database at high score of 908.0 and E-value of 0.0 PCR primers: FORWARD: M13 Forward BACKWARD: M13 Reverse Seq primer: T3 primer High quality sequence stop: 514 POLYA=No.
FEATURES	Location/Qualifiers
source	1..514 /organism="Bos taurus" /db_xref="taxon:9913" /clone_lib="Bovine Mixed Adipose cDNA library" /sex="Two males and one female mixed" /tissue_type="Adipose" /cell_type="Adipocyte" /dev_stage="Young adult" /lab_host="XLI-BlueMRF' strain" /note="Organ: Subcutaneous, Omental, Mesenteric adipose; Vector: uni-ZAPvPR; Site_1: EcoRI; Site_2: Xho I"
BASE COUNT	115 a 112 c 132 g 155 t
ORIGIN	
Query Match	60.0%; Score 15; DB 11; Length 514;
Best Local Similarity	100.0%; Fred. No. 4.2e+02;
Matches 15; Conservative	0; Mismatches 0; Indels 0; Gaps 0
OY	11 tgcacgccagcatc 25
Dn	242 TGCATGCCAGCATT 256
RESULT	30
LOCUS	A1957283
DEFINITION	u18sel2.x1 Susano mouse kidney mkia Mus musculus cDNA clone IMAGEF:2158894.3', mRNA sequence.
ACCESSION	A1957283
VERSION	A1957283.1 GI:5749992
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Mus. 1 (bases 1 to 524) Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,

Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurr, R., Ritter, E., Kohn, S., Shin, J., Jackson, T., Cardenas, M., McCann, R., Waterson, R. and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Maria M/washu-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@washington.wustl.edu
This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1002066
Seq primer: custom primer used
High quality sequence stop: 359.
Location/Qualifiers
1. .524
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone_image="2158894"
/clone_lib="Sugano mouse kidney mklia"
/sex="female"
/dev_stage="adult"
/lab_host="Dh10B"
/note="organ: kidney; Vector: pME18S-FL3; Site_1: DraIII (CACGCTG); Site_2: DraIII (CACCAATG); 1st strand cDNA was primed with an oligo(dT) primer
[ATGGGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCCCTACTGG], digested and cloned into distinct draIII sites of the pME18S-FL3 vector (5' site CACTGCTG, 3' site CACCAATG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTTAAGACTGCG and 3' end primer CGACCTCGACCTGCACCA."
BASE COUNT 151 a 102 c 147 g 124 t
ORIGIN
Query Match 60.0%; Score 15; DB 10; Length 524;
Best local Similarity 100.0%; Pred. NO. 4.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
QY 11 tgcacgcacgacatt 25
|||||
Db 352 tgcacgcacgacatt 338
RESULT 31
BG232022/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
536 bp mRNA EST 09-FEB-2001
naf3pe06.x1 Soares.NBMC Homo sapiens cDNA clone IMAGE:4143059 3'
similar to TR:P79434 P79434 CONNECTING PEPTIDE/TRANSMEMBRANE
/CYTOPLASMIC DOMAIN ;, mRNA sequence.
BG232022
BG232022.1 GI:12727177
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catartini; Homnidae; Homo.
1 (bases 1 to 536)
NCI-CCG http://www.ncbi.nlm.nih.gov/ncicgsp.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cga@pds-1@mail.nih.gov


```

/sex="Female"
/lab_host="DH10B"
/notes="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
EcORI; Site_2: EcORI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcORI and EcORI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcORI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      135 a      156 c      126 g      136 t
ORIGIN

Query Match      60.0%; Score 15; DB 13; Length 553;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 gcagcgtcgcgcga 20
Db      525 GCAGTGCATGCCA 511

RESULT 34
BF489544/c      566 bp      mRNA      EST      23-APR-2001
LOCUS
DEFINITION
AT25594.5prime AT Drosophila melanogaster adult testes pORH7
Drosophila melanogaster cDNA clone AT25594.5 similar to CG9085;
Fpan0009085 'protein kinase' located on: 3L.79E2-79E2:: 04/09/2001,
mRNA sequence.
BF489544
BF489544.2      GI:13755201
EST.
SOURCE
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 566)
Stapleton, M., Brokstein, P., Hong, L., Abmayan, A., Baxter, E., Berman
, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Partan
, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P.,
Liao, G., Miranda, A., Mista, S., Mungall, C.J., Nunoo, J., Pacleb, J.,
Paragas, V., Park, S., Phoumenavong, S., Wan, K., Yu, C., Lewis, S.E.,
Celisner, S. and Rubin, G.M.
BDGP/HMT AT Drosophila EST Project
Unpublished (2000)
On Dec 6, 2000 this sequence version replaced gi:11572845.
Contact: Stapleton, M.
BDCP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AE003597: arm:3L [22031266,22326831]
estimated-cyto:79D4-79E4: 04/09/2001
Plate: AT 255 row: H column: 10
High quality sequence stop: 527.
Location/Qualifiers
1..566
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="AT25594"
/clone_lib="AT Drosophila melanogaster adult testes pORH7"
/sex="male"
/dev-stage="0-3 day old Ore-R males"
/lab_host="Plates AT.10-AT.120: DHS-alpha. Plates
AT.121-AT.319: DHS-alpha Tona"
/notes="Organ: ADULT testes; Vector: pOTB7; Site_1: EcORI;
Site_2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles had dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into pOTB7. Plasmid cDNA library."

```

```

BASE COUNT      112 a      165 c      157 g      132 t
ORIGIN

Query Match      60.0%; Score 15; DB 11; Length 566;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 gcagcgtcgcgcgc 16
Db      511 GCAGCAGTGCATC 497

RESULT 35
C98142/c      567 bp      mRNA      EST      19-OCT-1998
LOCUS
DEFINITION
C98142 Rice callus Oryza sativa cDNA clone C0777_102, mRNA
sequence.
ACCESSION
C98142
C98142.1      GI:3760888
EST.
SOURCE
ORGANISM
Oryza sativa.
Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 567)
REFERENCE
AUTHORS
Sasaki, T. and Minobe, Y.
TITLE
Rice cDNA from callus
JOURNAL
Unpublished (1994)
CONTACT: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abr.affrc.go.jp. URL: http://rtp.dna.affrc.go.jp/
PROJECT="RGP"
Location/Qualifiers
1..567
/organism="Oryza sativa"
/strain="cultivar Nipponbare, sub-species Japonica"
/db_xref="taxon:4530"
/clone="C0777_102"
/clone_lib="Rice callus"
/notes="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
NotI. cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid. 1 others
of pBluescript II SK+ phagemid. 1 others
BASE COUNT      190 a      105 c      196 g      75 t
ORIGIN

Query Match      60.0%; Score 15; DB 11; Length 567;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      11 tgcacgcacacatt 25
Db      472 TGCAATGCCACGATT 458

RESULT 36
A2945591/c      586 bp      DNA      GSS      26-APR-2001
LOCUS
DEFINITION
2M0207H04F Mouse 10kb plasmid UUCG2M library Mus musculus genomic
clone UUCG2M0207H04 F, DNA sequence.
ACCESSION
A2945591
A2945591.1      GI:13809880
EST.
SOURCE
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;

```

REFERENCE 1 (bases 1 to 586)
 Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duvall, B., Hamill, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Kelly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.,
 and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10Kb
 plasmid inserts
 Unpublished (2000)
 JOURNAL COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0207 row: H column: 04
 Seq primer: GGTGTAAACGACCGCCACT
 Class: plasmid ends
 High quality sequence stop: 586.

FEATURES
 Source Location/Qualifiers
 1..586
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U06C2M0207H04"
 /clone_lib="Mouse 10Kb plasmid U06C2M library"
 /sex="Female"
 /lab_host="E. coli strain XL10-Gold, P1-resistant, F-"
 /note="Vector: PMD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (female) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adapted DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PMD42 (91147321419b/AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adapted mouse DNA was annealed to
 adapted vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 154 a 104 c 158 g 170 t

Query Match 60.0%; Score 15; DB 13; Length 586;
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 11 tgcatagcacgacat 25
 |||
 Db 189 tgcatagcacgacat 175

RESULT 37
 LOCUS AO783596 605 bp DNA GSS 03-AUG-1999
 DEFINITION HS_3121_A2_F07_T7A_C1T Approved Human Genomic Sperm Library D Homo
 sapiens genomic clone Plate-3121 Col-14 Row-K, DNA sequence.
 ACCESSION AO783596
 VERSION AO783596.1 GI:5691150
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 605)
 Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
 Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
 Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
 Hood, L.
 Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 JOURNAL COMMENT 99380589
 MEDLINE
 TITLE Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 3121 row: K column: 14
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 605.

FEATURES
 Source Location/Qualifiers
 1..605
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=3121 Col=14 Row=K"
 /clone_lib="C1T Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBeloBAC11; BAC clones in
 E-Coli DH10B"

BASE COUNT 187 a 187 c 123 g 92 t 16 others

Query Match 60.0%; Score 15; DB 13; Length 605;
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 10 gtgcatagcacgacat 24
 |||
 Db 248 gtgcatagcacgacat 234

RESULT 38
 LOCUS BG103212 623 bp mRNA EST 30-JAN-2001
 DEFINITION RH122_19_E10_g1_A003 Rhizome2 (RH122) sorghum prolingum cDNA, mRNA
 sequence.
 ACCESSION BG103212
 VERSION BG103212.1 GI:12618045
 KEYWORDS EST.
 SOURCE Sorghum prolingum.
 ORGANISM Sorghum prolingum.
 Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 623)
 Cordonnier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt
 L.H.
 An EST database from Sorghum: Sorghum prolingum rhizomes
 Unpublished (2000)
 JOURNAL COMMENT Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmp@pratt.uga.edu
 Seq primer: PolyTmix
 High quality sequence start: 48
 High quality sequence stop: 560

FEATURES
SOURCE

POLYA=NO.
Location/Qualifiers
1. .623
/organism="Sorghum propinquum"
/db.xref="taxon:132711"
/clone.lib="Rhizome2 (RHIZ2)"
/note="Organ: Rhizomes; Vector: Bluescript II from lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

BASE COUNT
171 a 140 c 149 g 163 t

ORIGIN

Query Match
Best Local Similarity 100.0%; Score 15; DB 11; Length 623;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 gcagtgcatgcgca 20
|||||
Db 394 GCAGTGCAATGCCA 408

RESULT 39
BG479445 730 bp mRNA EST 21-MAR-2001
LOCUS 603525920F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4649525 5',
DEFINITION mRNA sequence.
ACCESSION BG479445
VERSION BG479445.1 GI:13411724
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 730)
NIH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L10M1429 row: h column: 06
High quality sequence stop: 642.
Location/Qualifiers
1. .730
/organism="Homo sapiens"
/db.xref="taxon:9606"
/clone.lib="IMAGE:4649525"
/clone.lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10b (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(C). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using Zap-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT
169 a 184 c 193 g 184 t

ORIGIN

Query Match
Best Local Similarity 100.0%; Score 15; DB 11; Length 730;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ccgacgagtgatc 16
|||||
Db 706 CGCAGCAGTGCAATC 692

RESULT 40
BE194330 761 bp mRNA EST 02-MAR-2001
LOCUS HVSMH0085C19f Hordeum vulgare 5-45 DAP spike EST library
DEFINITION HVCDNA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMH0085C19f, mRNA sequence.
ACCESSION BE194330
VERSION BE194330.2 GI:13187270
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Hordeum.
1 (bases 1 to 761)
Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu, Y., Anderson, H., Dale, J., Henry, D., Kermode, S., Palmer, M., Ranbo, T., Saski, C., Schwartzbeck, J., Simmons, D., Choi, D.W., Main, D. and Wood, T.
Development of a genetically and physically anchored EST resource for barley genomics
Unpublished (2000)
On Jun 26, 2000 this sequence version replaced gi:8706516.
COMMENT
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Seq primer: AATTAAACCTTCACCTAAAGG
High quality sequence stop: 647.
Location/Qualifiers
1. .761
/organism="Hordeum vulgare"
/cultivar="Morex"
/db.xref="taxon:4513"
/clone="HVSMH0085C19f"
/clone.lib="HVSMH0085C19f"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI; For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley/> To order a clone see <http://www.genome.clemson.edu/orders>"

BASE COUNT
198 a 150 c 201 g 212 t

ORIGIN

Query Match
Best Local Similarity 100.0%; Score 15; DB 10; Length 761;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

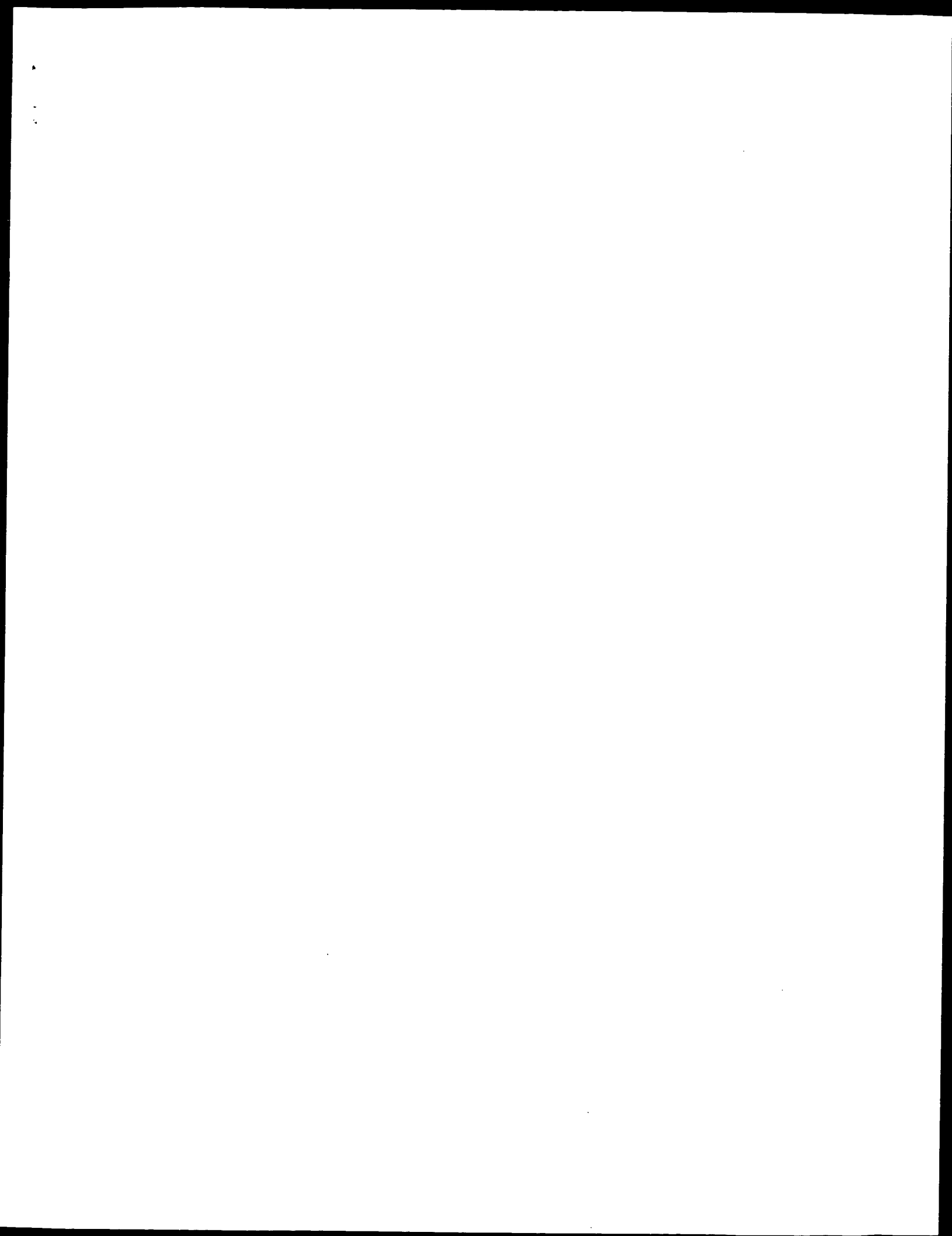
OY 10 gtcgacgacgacat 24
|||||
Db 54 GTGCATCCGCCAGCAT 40

RESULT 41
BF493796 767 bp mRNA EST 19-APR-2001
LOCUS AT02150.5prine AT Drosophila melanogaster adult testes P0787
DEFINITION Drosophila melanogaster cDNA clone AT02150 5 similar to CG9085; Fban0009085 'protein kinase' located on: 3L 79E2-79E2; 04/07/2001, mRNA sequence.
ACCESSION BF493796

VERSION BF493796.2 GI:13684306
 KEYWORDS EST.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 767)
 AUTHORS Stapleton, M., Brokstein, P., Hong, L., Aghayani, A., Baxter, E., Bernman, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan, D., Fitse, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Miranda, A., Misra, S., Munger, C.J., Nuno, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celiker, S., and Rubin, G.M.
 BDGP/HHMI AT Drosophila EST Project
 Unpublished (2000)
 On Dec 6, 2000 this sequence version replaced gi:11577097.
 COMMENT Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
 hit genomic AE003597: arm:3L [22031266,22326831]
 estimated-cyto:79D4-79E4: 04/07/2001
 Plate: AT 21 row: E column: 2
 High quality sequence stop: 661.
 Location/Qualifiers
 1..767
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="AT02150"
 /clone_lib="AT Drosophila melanogaster adult testes POT87"
 /sex="male"
 /dev_stage="0-3 day old Ore-R males"
 /lab_host="Plates AT 10-AT 120; DH5-alpha. Plates AT 121-AT 319; DH5-alpha Tona"
 /note="Organ: ADULT testes; Vector: POT87; Site_1: EcoRI; Site_2: XhoI; The mRNA for the testis library was made from testes and seminal vesicles hand dissected from 0-3 day old Ore-R males. RNA kindly provided by the lab of Margaret Fuller. Sized fractionated cDNAs were directly ligated into POT87. Plasmid cDNA library."
 BASE COUNT 166 a 203 c 206 g 192 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 11; Length 767;
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 cgcagcaggtgcac 16
 ||||||||||||||||
 Db 609 CGCAGCAGGTGCATC 595
 RESULT 42
 LOCUS BC299490/c 792 bp mRNA EST 21-FEB-2001
 DEFINITION HVSMEa0020F23f Hordeum vulgare seedling shoot EST library
 HVCDA0001 (Cold stress) Hordeum vulgare cDNA clone HVSMEa0020F23f, mRNA sequence.
 ACCESSION BC299490
 VERSION BC299490.1 GI:13086839
 KEYWORDS EST.
 SOURCE barley.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Hordeum.
 1 (bases 1 to 792)
 Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu, Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo

TITLE 'T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D.W., Main, D. and Wood, T.
 Development of a genetically and physically anchored EST resource for barley genomics
 JOURNAL Unpublished (2000)
 COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Seq primer: AATTACCCCTACTAAAGG
 High quality sequence start: 11
 High quality sequence stop: 774.
 Location/Qualifiers
 1..792
 /organism="Hordeum vulgare"
 /cultivar="Morex"
 /db_xref="taxon:4513"
 /clone="HVSMEa0020F23f"
 /clone_lib="Hordeum vulgare seedling shoot EST library
 HVCDA0001 (Cold stress)"
 /tissue_type="Seedling shoot"
 /lab_host="TJC121"
 /note="Vector: LambdaZAP; Site_1: EcoRI; Site_2: XhoI; For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley/> or order a clone see <http://www.genome.clemson.edu/orders>"
 BASE COUNT 123 a 229 c 235 g 183 t 22 others
 ORIGIN
 Query Match 60.0%; Score 15; DB 11; Length 792;
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 10 gtgcacgcagcat 24
 ||||||||||||||||
 Db 459 GTGCATCGCCAGCAT 445
 RESULT 43
 LOCUS BC309144 824 bp mRNA EST 22-FEB-2001
 DEFINITION HVSMEc004A20f Hordeum vulgare seedling shoot EST library
 HVCDA0003 (Etiolated and unstressed) Hordeum vulgare cDNA clone HVSMEc004A20f, mRNA sequence.
 ACCESSION BC309144
 VERSION BC309144.1 GI:13109991
 KEYWORDS EST.
 SOURCE barley.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Hordeum.
 1 (bases 1 to 824)
 Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu, Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo, T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D.W., Main, D. and Wood, T.
 Development of a genetically and physically anchored EST resource for barley genomics
 JOURNAL Unpublished (2000)
 COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Seq primer: AATTACCCCTACTAAAGG
 High quality sequence stop: 657.

FEATURES	source	location/Qualifiers
BASE COUNT	224 a 227 c 231 g 160 t	1..824
ORIGIN		/organism="Hordium vulgare" /culivar="Morex" /db_xref="taxon:4513" /clone="HVSMEC0004A20f" /clone_1ib="Hordium vulgare seedling shoot EST library HVCdNA0003 (Etiolated and unstressed)" /lissue_type="Seedling shoot" /lab_host="TTC121" /note="Vector: lambda2AP; Site_1: EcoRI; Site_2: XhoI; For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley/ To order a clone see http://www.genome.clemson.edu/orders "
BASE COUNT	247 a 177 c 253 g 147 t	
ORIGIN		
Query Match	60.0%; Score 15; DB 11; Length 824;	
Best Local Similarity	100.0%; Pred. No. 4.3e+02;	
Matches	15; Conservative 0; Mismatches 0; Indels 0; Caps 0;	
OY	1 ggcgcagcagtgcat 15 	
Db	801 GCGCAGCAGTGTCAT 815	
RESULT 44		
AJ282984		
LOCUS	AJ282984 843 bp mRNA EST 30-JUN-2000	
DEFINITION	4A3A-P2G5-R Anopheles gambiae immune competent 4A3A Anopheles	
ACCESSION	gambiae cDNA clone 4A3A-P2G5, mRNA sequence.	
VERSION	AJ282984	
KEYWORDS	AJ282984.1 GI:6930863	
SOURCE	EST.	
ORGANISM	African malaria mosquito.	
	Anopheles gambiae	
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
	Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae	
	; Anopheles.	
REFERENCE	1 (bases 1 to 843)	
AUTHORS	Dimopoulos,G., Casavant,T.L., Chang,S., Scheetz,T., Roberts,C., Donohue,M., Schultz,J., Benes,V., Bork,P., Ansoorge,W., Soares,M.B. and Kafatos,F.C.	
TITLE	Anopheles gambiae pilot gene discovery project: identification of mosquito innate immunity genes from expressed sequence tags generated from immune-competent cell lines	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (12), 6619-6624 (2000)	
MEDLINE	20300950	
COMMENT	Contact: Dimopoulos G Fotis C. Kafatos Laboratory European Molecular Biology Laboratory Meyerhofstrasse 1, 69117 Heidelberg, Germany. Location/Qualifiers 1..843	
FEATURES		
source	/organism="Anopheles gambiae" /strain="4A r/r" /db_xref="taxon:7165" /clone="4A3A-P2G5" /clone_1ib="Anopheles gambiae immune competent 4A3A" /cell_line="Immune competent 4A3A" /lab_host="E. coli DH10B" /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site1: EcoRI; Site_2: NotI; sequenced from forward priming site which reads from the 3' end of the cDNA. The 4A3A is a directionally cloned and normalized cDNA library that was constructed from the 4A3A cell line oligo-T primed cDNA according to: Bonaldo, Lennon & Soares (1996) : Normalization and Subtraction: Two approaches to Facilitate Gene Discovery" Genome Research 6, 791-806."	



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:51:46 : Search time 1757.07 Seconds
(without alignments)
234.726 Million cell updates/sec

Title: us-09-396-196f-6

Sequence: 1 gcaagtgatcgccagcattcgat 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 9

Total number of hits satisfying chosen parameters: 221953

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_hlgo_hum:*
31: em_hlgo_inv:*
32: em_hlgo_hum:*
33: em_hlg_hum:*
34: em_hlg_inv:*
35: em_hlg_rod:*
36: em_hlg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	1041	6	AR029499	AR029499 Sequence
2	25	100.0	1041	6	AR034916	AR034916 Sequence
3	25	100.0	1084	6	AL1530	AL1530 B10b gene o
4	25	100.0	1121	6	E00893	E00893 Genomic DNA
5	25	100.0	5793	1	E00810	J04423 E.coli 7.8-
6	25	100.0	5872	6	A38246	A38246 Sequence 1
7	25	100.0	5872	6	A38251	A38251 Sequence 6
8	25	100.0	5872	6	A93674	A93674 Sequence 1
9	25	100.0	5872	6	A93679	A93679 Sequence 6
10	25	100.0	5872	6	AR101809	AR101809 Sequence
11	25	100.0	5872	6	AR101810	AR101810 Sequence
12	25	100.0	11022	1	AE000180	AE000180 Escherich
13	25	100.0	13501	1	AE005258	AE005258 Escherich
14	25	100.0	297816	1	AP002553	AP002553 Escherich
15	19	76.0	3603	1	PSEHRPIA	L11382 Pseudomonas
16	18	72.0	2923	1	MAV250020	AJ250020 Mycobacte
17	18	72.0	5526	1	AF250776	AF250776 Unculture
18	17	68.0	479	11	G01446	G01446 Dm0466 Dros
19	17	68.0	10464	1	AE007981	AE007981 Agrobacte
20	17	68.0	53067	2	AC019940	AC019940 Drosophila
21	17	68.0	183937	2	AC037435	AC037435 Mus muscu
22	17	68.0	185200	3	AC092401	AC092401 Drosophila
23	17	68.0	191558	3	AC013431	AC013431 Drosophila
24	17	68.0	327446	3	AE003500	AE003500 Drosophila
25	16	64.0	1186	9	AF146614	AF146614 Erwinia c
26	16	64.0	4674	9	AK024397	AK024397 Homo sapi
27	16	64.0	11910	12	AF206717	AF206717 Shuttle v
28	16	64.0	49587	3	DMB448C10	AL133503 Drosophila
29	16	64.0	60573	2	AC018267	AL135267 Drosophila
30	16	64.0	82289	8	ATAC009895	AC009895 Arabidops
31	16	64.0	117968	8	HS272E8	293929 Human DNA s
32	16	64.0	140825	8	AP002817	AP002817 Oryza sat
33	16	64.0	146081	8	AP001366	AP001366 Oryza sat
34	16	64.0	154306	2	AC034114	AC034114 Homo sapi
35	16	64.0	177380	9	AC002457	AC002457 Homo sapi
36	16	64.0	180700	2	AC084136	AC084136 Homo sapi
37	16	64.0	183439	3	AC007809	AC007809 Drosophila
38	16	64.0	232744	3	AE003705	AE003705 Drosophila
39	16	64.0	237119	2	AC017740	AC017740 Drosophila
40	16	64.0	289090	3	AE003424	AE003424 Drosophila
41	15	60.0	1185	8	AF374832	AF374832 Saxifraga
42	15	60.0	1389	8	NTPPLG	X61102 N.tabacum g
43	15	60.0	1713	8	NTPPLG	X61102 N.tabacum g
44	15	60.0	1907	9	HSSYNTRO5	AJ003029 Homo sapi
45	15	60.0	2616	9	AB048915	AB048915 Macaca fa
46	15	60.0	3369	8	AF274670	AF274670 Tetraodon
47	15	60.0	3369	8	AF274671	AF274671 Trochodon
48	15	60.0	3677	1	PSEIRAMH	M11035 P.syringae
49	15	60.0	5618	9	AB028980	AB028980 Homo sapi
50	15	60.0	10866	1	AE008017	AE008017 Agrobacte
51	15	60.0	10978	1	AE004795	AE004795 Pseudomon
52	15	60.0	11684	1	AE007551	AE007551 Clostridi
53	15	60.0	14846	1	AE007050	AE007050 Mycobacte
54	15	60.0	15916	1	MTCY9C4	Z77250 Mycobacte
55	15	60.0	17988	1	AE007098	AE007098 Mycobacte
56	15	60.0	23873	2	AC018333	AC018333 Drosophila
57	15	60.0	35209	1	AF065159	AF065159 Bradyrhiz
58	15	60.0	44201	1	MTCY180	Z97193 Mycobacteri
59	15	60.0	57237	3	CEY62E10A	AL132865 Caenorhab
60	15	60.0	80332	2	AC010038	AC010038 Drosophila
61	15	60.0	86217	9	HSU92032	U92032 Homo sapien
62	15	60.0	86896	9	HS93N13	Z84489 Human DNA s
63	15	60.0	94814	8	AB046426	AB046426 Arabidops
64	15	60.0	94887	8	AC012394	AC012394 Arabidops
65	15	60.0	100806	8	AC015450	AC015450 Arabidops
66	15	60.0	110000	2	LMFLCHR34-06	Continuation (7 of
67	15	60.0	110000	2	LMFLCHR34-07	Continuation (8 of
68	15	60.0	117335	2	AL158848	AL158848 Homo sapi
69	15	60.0	120206	9	HS439F8	AL021392 Human DNA
70	15	60.0	129949	2	AC022940	AC022940 Homo sapi

```

71 15 60.0 134213 9 CNS05TEV
72 15 60.0 134213 9 CNS05TEV
73 15 60.0 147223 2 AC090713
74 15 60.0 168674 2 AC025645
75 15 60.0 169524 2 AC023111
76 15 60.0 181861 2 AC016322
77 15 60.0 182080 3 AC016322
78 15 60.0 183521 2 AC091609
79 15 60.0 194781 2 AC084279
80 15 60.0 197303 2 AC012506
81 15 60.0 201569 2 HS11815
82 15 60.0 207842 9 AC010276
83 15 60.0 213732 1 AC001862
84 15 60.0 242407 2 AL589790
85 15 60.0 256010 3 AC003657
86 15 60.0 294942 3 AC011662
87 14 56.0 3338 8 HV0234505
88 14 56.0 363 1 AB026474
89 14 56.0 367 1 G24898
90 14 56.0 372 1 AB026469
91 14 56.0 384 1 AB026470
92 14 56.0 390 1 AB026475
93 14 56.0 399 1 AB026458
94 14 56.0 401 1 G29590
95 14 56.0 426 1 AB026457
96 14 56.0 432 1 AB026455
97 14 56.0 435 1 AB026460
98 14 56.0 450 1 AB026456
99 14 56.0 574 2 AC031725
100 14 56.0 606 2 AC056445

```

ALIGNMENTS

```

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
  1 (bases 1 to 1041)
  Patton,D.Andrew.
  Enhanced biotin biosynthesis in plant tissue
  JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
  FEATURES
    Source
      1..1041
      /organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.00038;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 gcaggtgcatgcagcattcgat 25
    |||
Db 84 GCAGGTGCATGCCACGATTTCGAT 108

```

```

RESULT 2
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.

```

```

ORGANISM Unknown.
REFERENCE
  1 (bases 1 to 1041)
  Patton,D.A.
  Transgenic plants having increased biotin content
  JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
  FEATURES
    Source
      1..1041
      /organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.00038;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 gcaggtgcatgcagcattcgat 25
    |||
Db 84 GCAGGTGCATGCCACGATTTCGAT 108

```

```

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION B10B gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE
  1 (bases 1 to 1084)
  Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
  Escherichia.
  JOURNAL Patent: GB 2216530-A 16 11-OCT-1989;
  FEATURES
    source
      1..1084
      /organism="Escherichia coli"
      /db_xref="taxon:562"
      /gene="B10B"
      /gene="BIOB"
      /gene="BIOB"
      /codon_start=1
      /transl_table=11
      /protein_id="CAA00965.1"
      /db_xref="GI:490219"
      /db_xref="SWISS-PROT:P12996"
      /translation="MAHRPRMTLSQVTELPFKPLIDLLFEAQYHROHDFPROVST
      LLSIKTGAQPEPCCKYCPOSSRYKGLAEERLMEVROYEESRRKKAAGSTRFGAMW
      KPHPRDMPYLEQYQVYKAKGLEACHTLGLTSESQVRLNAGIDYVNHLDLSPER
      VGNITITRYOEIRDTLEKVRDAGIKVSGSIVGIGETVQKDRAGILLQLANLPPPS
      VPIINLVKVGTPPLADNDVDADFIRTAIVARIIMPTSVYRLSAGREOMNEQOAMC
      FMAGANSIFYGCKLITTPNPEEDKDLQIFRRLGLNPOOTAVLADNEQQQLTQALMT
      PTDDEYVNAAL"

```

```

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.00038;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 gcaggtgcatgcagcattcgat 25
    |||
Db 107 GCAGGTGCATGCCACGATTTCGAT 131

```

```

RESULT 4
LOCUS E00893 1121 bp DNA
DEFINITION Genomic DNA encoding biotin Synthetase.
E00893

```

```

ACCESSION      E00893
VERSION        E00893.1
KEYWORDS       GI:2169154
SOURCE         JP 1986149091-A/1.
ORGANISM       Escherichia coli.
                Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
                Escherichia.
REFERENCE      1 (bases 1 to 1121)
AUTHORS       Hirano, Y., Kojima, T. and Kimura, H.
TITLE         DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
                PRODUCTION OF BIOTIN
JOURNAL       Patent: JP 1986149091-A 1 07-JUL-1986;
COMMENT       NIPPON SODA CO LTD
                OS Escherichia coli
                PN JP 1986149091-A/1
                PD 07-JUL-1986
                PF 24-DEC-1984 JP 1984272605
                PI HIROMO YOSHIMITSU, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
                CI2N15/00,CI2N1/20,CI2P13/18,(CI2N1/20,CI2P1:19),(CI2P13/18,PC
                CI2P1:19);
                CC strandedness: Double;
                CC topology: Linear;
                CC hypothetical: No;
                CC anti-sense: No;
                CC *source: strain=Escherichia coli NS101;
                CC feature is identified by experimental;
                FH key Location/Qualifiers
                FT CDS 42..1079
                /product="biotin synthetase".
FEATURES       source
                location/Qualifiers
                1..1121
                /organism="Escherichia coli"
                /db_xref="taxon:562"
BASE COUNT    289 a 296 c 325 g 211 t
ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 1121:
Best Local Similarity 100.0%; Pred. No. 0.00038;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcaggtgcacgcacacattcgat 25
    |||||||||||||||||||
Db 125 GCAGTGCATCGCCAGCATTCGAT 149
RESULT 5
ECOBIO 5793 bp DNA BCT 28-FEB-1994
LOCUS      E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
DEFINITION (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioC
            protein, and dehydrobiotin synthetase (bioD), complete cds.
ACCESSION  J04423
VERSION    J04423.1
KEYWORDS   GI:145422
SOURCE     Escherichia coli (strain K-12) DNA.
ORGANISM   Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE  1 (bases 1 to 5793)
AUTHORS    Otsuka,A.J., Buonocristiani,M.R., Howard,P.K., Flamm,J. and
            Johnson,O.
TITLE      The Escherichia coli biotin biosynthetic enzyme sequences
JOURNAL    J. Biol. Chem. 263, 19577-19585 (1988)
MEDLINE    89066784
COMMENT    Draft entry and computer-readable sequence [1] kindly submitted by
            A.Otsuka, 09-NOV-1988.
FEATURES   location/Qualifiers
                source
                1..5793
                /organism="Escherichia coli"
                /strain="K-12"
                /db_xref="taxon:562"
                complement(98..574)
                /note="ORF 1"
                /codon_start=1
                /transl_table=1
                /protein_id="AAA23513.1"
                /db_xref="GI:451168"
                /translation="MKLISNDLRDGDLPKPHRVENGMDGDNISPHLAMDVPAGTK
                SYVCVCPDADAPGSGMHHVVVNLNLPADRVLPQGGSLVAMPDGVQTRDECKTG
                YDCAAPKGETHRVITFYALDIERLDVDEGASAAVGVNHFHSLASITAMS"
                complement(633..1925)
                /gene="bioA"
                complement(633..1925)
                /gene="bioA"
                /codon_start=1
                /transl_table=1
                /product="7,8-diamino-pelargonic acid aminotransferase"
                /protein_id="AAA23514.1"
                /db_xref="GI:457106"
                /translation="MTTDLAPDQRIHTPTSMSPLPYVVSAGECELLISGR
                LVDGSSWNAALIHGVNHPQLNAAKSOIDAMSHVMFGITPALELCKRLVAMSGRN
                ALECVPLADSGSAVEVAMKMAIQWQAGARORLTPRNGYHGDTEGAMSCPDN
                GMSLWKGYLPENLFAPOSMDGEMDERMDVGFARLAAHRIEALVILPIYOGA
                CWRMYHPENLKRIRKICDREGILLADELFGPTGKLPFAGEHEIALPDILCGKA
                LTGTTTLSTLTTRVAVETISNGEACGCPMPTPEWGNLACAAANASLALIESGDM
                QVADIEVOLREQLAPARDENAVDVGALGVYETTPYVMAALQKFEVQGWIR
                PFGLIYLPMPYILPQOLRLTAAVNRAVODETFECQ"
                2012..3052
                /gene="bioB"
                2012..3052
                /gene="bioB"
                /codon_start=1
                /transl_table=1
                /product="biotin synthetase"
                /protein_id="AAA23515.1"
                /db_xref="GI:145425"
                /translation="MAHRPRTLSQVTELPFKPLDLLPEAOVHRQHPDPOVST
                LLSKTGACPEDCKYCQPSRYRTGLEAEKLEMEVYVLEBARKAKAGSTRCMGAM
                KNHPRDMPYLEQWVOCVKAMGLEACWTLTSLSEQORLANGLDYTNHNLDTSPF
                YGNIITRTYQERLDLEKRVADGIRKCSGIVGLGETVYDRAGLLTLQNLPTPES
                VPINMLVVKVGTPLADNDVDADPFIETIAVARIMPTSVRLSASREGDNEGTOMC
                PMGANSIFPGCKLTLTPNEBEDKDLQLFKLCINPQOTAVLAGDNEQOORLEQALMT
                PTDDEYVNAAL"
                3049..4203
                /gene="bioC"
                3049..4203
                /gene="bioC"
                /codon_start=1
                /transl_table=1
                /product="7-keto-8-amino pelargonic acid synthetase"
                /protein_id="AAA23516.1"
                /db_xref="GI:145426"
                /translation="MSWQEKINAAIDARRADALRRRYPYAGAGRWLVADRDQYINF
                SSNIYIGISHHPOTIRAMOGAEQOFGSGSGSGSYVHQALEBELLEWGYSKRA
                LLFTSGRANAOVIAAMAKEDRIAAORLSHASLSLEASLSPSLRREFAHNDYHAR
                LLASPCPEQOMVNTIEGVFSMDGDSAPLAETQVYQOINGLMNDADHGTGVEOEG
                SCMLQAKVPELVVTEFGKRGVSAAVLCSTVADYLLQFRRLITSTSPRPAQDAL
                RASLAVIRSDGDAVRREKLAALITRFAGVODLPETTLADSCAIOPLIVDNSRAQDL
                AEKLROGCWTAIRPPTVPAGTARLRLTLAAHEMODIDRLLEVLHGNC"
                4190..4945
                /gene="bioC"
                4190..4945
                /gene="bioC"
                /codon_start=1
                /transl_table=1
                /protein_id="AAA23517.1"
                /db_xref="GI:145427"
                /translation="MATYNKQAITAAGRAAHYEHQADIDORSADALLAMPORKYT
                HLVDGCGPQWMSRHRNRHNOVYRALDISPMILVQAKQKDAHRYLAGDIESIPIATA
                TFDLAWSNLAVAWCGNLSTALRELRYRVRKGVAVATTVLVOGSLPERHQANVADERP

```



```

stem_loop      5583..5605
terminator     5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT    1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00037;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 gcaggtgcacgcacgacatccgat 25
        |||
Db      200 GCAGGTGCATCCGACGATTCGAT 224

RESULT 7
A38251      5872 bp      DNA      PAT      05-MAR-1997
LOCUS      A38251
DEFINITION Sequence 6 from Patent WO9408023.
ACCESSION  A38251
VERSION     A38251.1 GI:2294849
KEYWORDS
SOURCE      Escherichia coli.
            Escherichia coli.
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL     Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 9600228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 85016947 960227.
            Location/Qualifiers
              1..5872
                /organism="Escherichia coli"
                /strain="DSM498"
                /db_xref="taxon:562"
                /clone="PBO30A15-9"
                /standard_name="BIOF RBS"
                /gene="BIOF"
                /gene="BIOF"
                /standard_name="8-AMINO-7-OXONONANATE SYNTHASE"
                /EC_number="2.3.1.47"
                /codon_start=1
                /transl_table=11
                /number=2
                /evidence=experimental
                /product="KAPA SYNTHASE"
                /protein_id="CAA02329.1"
                /db_xref="GI:2294850"
                /translation="MSMOKETNALADARRADARRRYPAAGAGRMLVADDDROYLNF
                SSMYDGLSHHPOTITRAMOGAGQFGSCSGSHVGY SVHQALEELAEMLYSRA
                LFLTSGFANQAVTAAAMAKEDRIADRLSHSLERASLSPSQLRRFAINDYTHLAR
                LLASPCGQOMVTEGVFSMDGSDAPLAETIQVYQHNGMLWDDAHGTGTEGGRG
                SCWLQVKPELVITFGKGFVSGAAVLCSTVADYLQFARHLIYSTSPPAQDAL
                RASLAVTRSDGQARRREKLAALITRFAGVQDDPFTLADSCSAIOLIVGDNRAIOL
                AERLRQGCWVTALRPPTVPAGTARLRLTLTAHFMQDIDRLLEVLHGNC"
RBS        3030..3045
            /standard_name="BIOB RBS"
            3043..3753
            /gene="BIOB"
            /gene="BIOB"

```

```

CDS
3043..3753
/gene="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/translation="MSKRRFYVGTDTENVKTYASCALLOAKAAGRTAGYKPVASGS
EKTPGELRNSDALAQNRSSLDIATVATVPYFAEPTSHITISADGRPIESLVSAG
LRALBQADWVLVEGAGWFTPLSDTFPADWTOEQLPVLLVGVKLGKINHAULTA
OVIQAHGGLTLAGVANDVTPPGKRHAKEYMTTLTRMIPALIGEIPMLAENPENATGK
YINLAVDASTIGFTSRL"
BASE COUNT    1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00037;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 gcaggtgcacgcacgacatccgat 25
        |||
Db      200 GCAGGTGCATCCGACGATTCGAT 224

RESULT 8
A93674      5872 bp      DNA      PAT      22-JAN-2000
LOCUS      A93674
DEFINITION Sequence 1 from Patent EP0798384.
ACCESSION  A93674
VERSION     A93674.1 GI:6741862
KEYWORDS
SOURCE      Escherichia coli.
            Escherichia coli.
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch,O. and Brass,J.
TITLE      Biotechnological method of producing biotin
JOURNAL     Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
COMMENT     Location/Qualifiers
              1..5872
                /organism="Escherichia coli"
                /strain="DSM498"
                /db_xref="taxon:562"
                /clone="PBO30A-15/9"
                /function="PROMOTER PTAC"
                /evidence=experimental
                /standard_name="PROMOTER PTAC"
                /evidence=experimental
                /standard_name="BIOB RBS NO.9"
                /evidence=experimental
                /gene="BIOB"
                /gene="BIOB"
                /codon_start=1
                /transl_table=11
                /evidence=experimental
                /product="BIOTIN SYNTHASE"
                /protein_id="CAB69590.1"
                /db_xref="GI:6741863"
                /translation="MAHRPRMWTLSQVTELFKPKLLDLFEAQOVVHROHFDPROVOST
                LLSIKTGACPEDCKYCKPOSSRYKKTGLEARLMEVQVLESARKAKAGSTRCMGAAM

```

KNPEDMPEYLBQWGVKANGLEACMTIGTSESQACRLANAGIDYNNHNDJSPF
YGNITITRTYOEKDLLEKVRDAGIKVCGGIVGGETVKDAGLLDLANLPTPEP
VPLNMLVKVKGKPLADNDVDADFRTITAVARIMKSYVRLSAGRQMNQOTQAMC
FNAGANSIFYGCKLITTPNPEEDKDLQLEFRKILGLNPQCAVLACINDEQOQRLQGLMT
PIDEYYNAAL"

RBS
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"

gene
2295..3050
/translation="MATYKQATAAAGRAAHYEOHADLROSADALLAMLPKRYT
HYLDAGCGGMMWRHREHRAQVATLDSIPMLYGRKDAADHYADGIESPLATA
TPDLAMSLAYOMCGNLSTALRELYRVRPKGVAFPTLNGSLDELHQAQVADERP
HANRFLPDELEIOSINGVXOHHIDPITLNTDDALSNRSLKIGIGHVHLEGRDPRIL
TRSOLOLQIAMPQOGKPLTYHLEGLVARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE: 8-AMINO-7-
OXONONANATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"

gene
5038..5100
/translation="MTTDDLAFDORHIMPTYSPLPYVVSAGCELLSDGR
LVDSMSWMAIHGYNHNPOLNAARKSOIDAKSHFGITTHAPATELCRIYAMTPRP
LECVLADSGVAEVAAMKALOWAKGAEKROPLFRNGYHDTGAMSVCDPDS
MHSLWGLPENLEFAPAPQSMGEMDERMVGARLMAAHRHETAVIIEPIVQAG
GGRWHPDEM.KRIKICDRBILITADELITGFRIGKLEACEHAETAPDILIGAL
TGGTMTSATITITREVAETTSNAGCFMGGPFMGNPLACAAANSLAPDILIGAL
QVADIEVQLREQLAPARDAEMVADVRLGAGIVETTHPVNMAALOKFEVEQVWIRP
FGKLIIMPYIILPQOLRLTAANRAVDDEFQO"
5088..5100
/standard_name="ORF1 RBS"
5098..5574
/gene="ORF1"
5098..5574
/gene="ORF1"
5098..5574
/gene="ORF1"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"

gene
5583..5605
/translation="MKLISNDLRDGLPLRHVNGMGYDGNISPHLAMDVPACT
SEVVCYDDPAPGSGMMHVVYVNLPAIDRVLPDQSGSLVAMPDGVLOTTRDFGTG
YDCAAPKGETTHYIFTVALDIERIDVDEGASGAVGNNHSHSLASTAMS"
5583..5605
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
1318 a 1552 c 1695 g 1307 t

BASE COUNT
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00037;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcaagtgcacccagacattccat 25
|||||
Db 200 GCAGGTGCATCCGCCAGCATTTGGAT 224

RESULT 9
A93679
LOCUS A93679 5872 bp DNA
DEFINITION Sequence 6 from Patent EP0798384.
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE
1 (bases 1 to 5872)
AUTHORS Birch, O. and Brass, J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)

FEATURES
source
location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"

RBS
3030..3045
/translation="MSMOEKINAAADARRADALRRYPVAGAGRWLVADROYLNF
SNDYGLSHRQITRAOQGAEOFGTSGSGSHVSGSVVHQALEELALBWLVSRA
LUTGFRANQAVIAMAKEDRIAADRLSHASLLEASLSPSQLRRFAHVDYHILAR
LLASPCQGOVWTEGVSMQDSAPLAEIDQVIOQHNGMLVDAHGTGIGGOGG
SCMIOKVPBELIVTFEGKFGVSGAVALCSSVADYLLQFARHLIYSNMPADAOAL
RSLAVIRSDDEGDAREKLAALITFRAGVODLPPTLADSCSAIPLIVGNSRALDL
AEKLRQCGWTAIRPTVPVAGTARLRLITLTAHEMODIRLLEVLHONG"
3030..3045
/standard_name="BIOD RBS"
3043..3753
/gene="BIOD"
3043..3753
/gene="BIOD"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"

gene
3043..3753
/translation="MSKRYEYTGTDTEVGKTVASCALLQAAKAGYRTAGYKPVASGS
EKTPGELNSDALALORNSSLDIYATVAVNPTAEPTSPHIIISOEGRPJESLVMASG
LRALQOQADWLVVAGAGGWTPLSDTFEPADWVTOEDLPVILVGVGLGGINAMILTA
QVIOHAGLTLAGWANDVTPPGKRHAHYMTTLRMIPARPLIGELPMLAENPEANAATGK
YINLAFVASTLGTFSRL"

BASE COUNT
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00037;

```

Matches      25; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

Oy      1      gcaggtgcatgcgcagcatttcgat 25
      |||||||
Db      200      GCAGGTGCATGCCAGCATTTTCGAT 224

RESULT      10
LOCUS      AR101809      5872 bp      DNA
DEFINITION      Sequence 1 from patent US 6083712.
ACCESSION      AR101809
VERSION      AR101809.1      GI:12812607
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unclassified.
REFERENCE      1 (bases 1 to 5872)
AUTHORS      Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE      Biotechnological method of producing biotin
JOURNAL      Patent: US 6083712-A 1 04-JUL-2000;
FEATURES
SOURCE      1. 5872
              /organism="unknown"
BASE COUNT      1318 a 1552 c 1655 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00037;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1      gcaggtgcatgcgcagcatttcgat 25
      |||||||
Db      200      GCAGGTGCATGCCAGCATTTTCGAT 224

RESULT      11
LOCUS      AR101810      5872 bp      DNA
DEFINITION      Sequence 6 from patent US 6083712.
ACCESSION      AR101810
VERSION      AR101810.1      GI:12812608
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unclassified.
REFERENCE      1 (bases 1 to 5872)
AUTHORS      Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE      Biotechnological method of producing biotin
JOURNAL      Patent: US 6083712-A 6 04-JUL-2000;
FEATURES
SOURCE      1. 5872
              /organism="unknown"
BASE COUNT      1318 a 1552 c 1655 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00037;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1      gcaggtgcatgcgcagcatttcgat 25
      |||||||
Db      200      GCAGGTGCATGCCAGCATTTTCGAT 224

RESULT      12
LOCUS      AE000180      11022 bp      DNA
DEFINITION      Escherichia coli K12 MG1655 section 70 of 400 of the complete
genome.
ACCESSION      AE000180      U00096

```

```

VERSION      AE000180.1      GI:1786988
KEYWORDS
SOURCE      Escherichia coli K12.
ORGANISM      Escherichia coli K12
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE      1 (bases 1 to 11022)
AUTHORS      Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
Mau, B., and Shao, Y.
TITLE      The complete genome sequence of Escherichia coli K-12
JOURNAL      Science 277 (5331), 1453-1474 (1997)
MEDLINE      92745617
PUBMED      9278503
REFERENCE      2 (bases 1 to 11022)
AUTHORS      Blattner, F.R.
TITLE      Direct Submission
JOURNAL      Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
3 (bases 1 to 11022)
AUTHORS      Blattner, F.R.
TITLE      Direct Submission
JOURNAL      Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
4 (bases 1 to 11022)
AUTHORS      Plunkett, G. III.
TITLE      Direct Submission
JOURNAL      Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Fredrick R. Blattner director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using GenMark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
30332 (e-mail: mark@amherst.gatech.edu). Open reading frames that
have been correlated with genetic loci are being annotated with CG
Site Nos., unique ID nos. for the genes in the E. coli Genetic
Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlyn. A public version of the database is accessible
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). ** The E. coli K12 sequence and
its annotations are periodically updated; this is version M54. No
sequence changes. Annotation updates: updated gene identifications
and products; all new functional assignments courtesy of Monica
Riley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'b' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of
labels. This should allow them to be searched for in Entrez as gene
names.
FEATURES
source      Location/Qualifiers
              1..11022
              /organism="Escherichia coli K12"
              /strain="K12"
              /sub_strain="MG1655"
              /db_xref="taxon:83333"
              <1..130
              /note="REP (repetitive extragenic palindromic) element;
              contains 4 REP sequences"
              complement(147..1430)
              /gene="ybhC"
              /note="b0772"

```

```

CDS
    complement(147..1430)
    /gene="ybhC"
    /function="orf; Not classified"
    /note="t427; 98 pct identical to fragment YBH_CECOLI
    SW:P46130 (300 aa) but contains 127 additional C-terminal
    residues"
    /codon_start=1
    /transl_table=11
    /product="putative pectinesterase"
    /protein_id="AAC73859.1"
    /db_xref="GI:1786989"
    /translation="MNTFSVRLALALAFVTLTACSSPTPPDPSDOTAGTSRPI
    LSAKEQNDAQHYFASLTSPGAAMNPSITLPADDFVAGPGAGTQVHTHTIOAAVD
    AAIIRKTRKROYIAVPGEGYGVVPPAPAGSTITLYGCEKRPDKITSLDGGHSPA
    DMRDYNGPKYMGPKPAMYMDSCOSKRSITGYCSAVENSONNGLOLQNTLENT
    LGDSVDAGNHPVALRTDGOVQINNVNITLGRONTFFVYNSGVNQLQNTLENT
    NSYIEGDVIVSGKAVVEFNTEFRVNSRQGEAYVAPATLSNLYGFLAVNSKFN
    AFQGVADQLGRSLVDVANTNGOVITRDSAINCEGFYAKPMADAVISNRPFGNNGSVD
    DMDEIQRLNDINRYNRMMEYNRNGVSKVAEAKK"
    complement(1507..1535)
    /note="factor Sigma70; predicted +1 start at 806574"
    complement(1582..2058)
    /gene="ybhB"
    /note="b0773"
    complement(1582..2058)
    /gene="ybhB"
    /function="orf; Unknown"
    /note="t158; 99 pct identical to YBH_B-ECOLI SW: P12994"
    /codon_start=1
    /transl_table=11
    /product="orf, hypothetical protein"
    /protein_id="AAC73860.1"
    /db_xref="GI:1786990"
    /translation="MKLISNDLRGSDKLRHVRNGKGYDQDNISPHLAWDVPAQTK
    SEVYVCGPDAPRTGSGMHWVYVNIADTRVLPQGFSGLVAMPDGLQTPDQKTC
    YDGAAPKPGETHRTITVHALDIERIDDEGASGAMGFWHFSLASATYMFs"
    complement(2108..2124)
    /note="central position to predicted promoter:85.5"
    /bound_moiety="Rhas predicted site"
    complement(2117..3406)
    /gene="b10A"
    /note="b0774"
    complement(2117..3406)
    /gene="b10A"
    /EC_number="2.6.1.62"
    /function="enzyme; Biosynthesis of cofactors, carriers:
    Biotin"
    /note="t429; 100 pct identical to B10A-ECOLI SW: P12995"
    /codon_start=1
    /transl_table=11
    /product="7,8-diaminopelargonic acid synthetase"
    /protein_id="AAC73861.1"
    /db_xref="GI:1786991"
    /translation="MTMDLAFDQRIHMPYTSMTSPFLPYVVSAGCELLISDCRR
    LVDSSSWMAIHGYNHNPOLNAAMKSOIDKAMSHVQRTFTHAETLCRLVAMTPQP
    LECYFLADSSSAVEVAMKVALQYWKQAGARQFRTFRNGYQDTEFGASVCDPNS
    GMSLWKGYLPENLFAPAQSRMDGEMDERMVFARMAHREHIAVITEPTVQAG
    GMRVYHPWELKRIKICDRGILLADEIATGFRGKIFACCHAEIAPILLTGKAI
    TGGTMTLSATLTREVAETISGEAGCFMGPEMCPPLACAANASIALTESGDMCO
    QVADIEVOLREOLAPARDAEMVAVRVLGIVGVEPTTHPVNMAALOKFEVQGVWIRP
    FGKLIYMPYIILPOOLRLTAANRAVDETFECO"
    complement(2193..2221)
    /gene="b10A"
    /note="factor Sigma70; predicted +1 start at 807260"
    3411..3450
    /note="central position to biob promoter: -20"
    /bound_moiety="Biob documented site"
    3411..3450
    /note="central position to predicted promoter: -20; genetic
    evidence for the site"
    /bound_moiety="Biob predicted site"
    complement(3411..3450)
    /note="central position to predicted promoter:50"

protein_bind
    /bound_moiety="Biob predicted site"
    complement(3411..3450)
    /note="central position to biob promoter:50"
    /bound_moiety="Biob documented site"
    3413..3441
    /note="factor Sigma70; promoter biob; documented +1 at
    808525"
    complement(3447..3473)
    /note="factor Sigma70; promoter biob; documented +1 at
    808515"
    3493..4533
    /gene="b10B"
    /note="b0775"
    3493..4533
    /gene="b10B"
    /EC_number="2.8.1.-"
    /function="enzyme; Biosynthesis of cofactors, carriers:
    Biotin"
    /note="t046; 99 pct identical to B10B-ECOLI SW: P12996"
    /codon_start=1
    /transl_table=11
    /product="biotin synthesis, sulfur insertion?"
    /protein_id="AAC73862.1"
    /db_xref="GI:1786992"
    /translation="MAHREPWTLISQVTELEKPLDLLEFAQOVHROHFDPROVST
    LLSIKTGACPEDECKYCPOSSRYKTGLEARLVEQVLEASARKAKAAGSTRGMAW
    KNHERDMPYLEQMOVGVKAMGLEACMTIGTISESOARLANAGLIDYVHNLDTSPEF
    YGNIITTRTYOERLDLTKERYRDAGTKVSGGIVGCEYKDRAGLILLOLANPTPES
    VPINMLVKYKGPRLADNDVDADPDRITAVARIMPTSYKLSKREDMNQTAMC
    PMGANSITVCGCKLLTPNPEEDKQLQIRKGLINPOQAVIAGNEQOORLEQALMT
    PVDERYNAAL"
    4530..5684
    /gene="b10F"
    /note="b0776"
    4530..5684
    /gene="b10F"
    /EC_number="2.3.1.47"
    /function="enzyme; Biosynthesis of cofactors, carriers:
    Biotin"
    /note="t0384; 100 pct identical to B10F-ECOLI SW: P12998"
    /codon_start=1
    /transl_table=11
    /product="8-oxo-7-oxononanoate synthase"
    /protein_id="AAC73863.1"
    /db_xref="GI:1786993"
    /translation="MSWQEKINAALDARRAADALRRRYVPAQAGRWLVADROYLNF
    SSNDYLGISHPOIIRAMOGCAEFGICSGSGHVSQSVHQALEELFARMIGYRA
    LFIISGFANQAVIAAMAKEDRIADRLSHASILEASISPSORRRAHNDYHILAR
    LASPCGGQWVTEGVFSMDCSDAPLAELIOQYTOOHNGWMLVDDAHMETGYIGDGRG
    SCWLDKVRKELLVTFGKRGVSGAAVLCSSIVADYLLQFARHLIYSTSMPPAOAOL

gene
    4530..5684
    /gene="b10F"
    /note="b0776"
    4530..5684
    /gene="b10F"
    /EC_number="2.3.1.47"
    /function="enzyme; Biosynthesis of cofactors, carriers:
    Biotin"
    /note="t0384; 100 pct identical to B10F-ECOLI SW: P12998"
    /codon_start=1
    /transl_table=11
    /product="8-oxo-7-oxononanoate synthase"
    /protein_id="AAC73863.1"
    /db_xref="GI:1786993"
    /translation="MSWQEKINAALDARRAADALRRRYVPAQAGRWLVADROYLNF
    SSNDYLGISHPOIIRAMOGCAEFGICSGSGHVSQSVHQALEELFARMIGYRA
    LFIISGFANQAVIAAMAKEDRIADRLSHASILEASISPSORRRAHNDYHILAR
    LASPCGGQWVTEGVFSMDCSDAPLAELIOQYTOOHNGWMLVDDAHMETGYIGDGRG
    SCWLDKVRKELLVTFGKRGVSGAAVLCSSIVADYLLQFARHLIYSTSMPPAOAOL

CDS
    complement(2117..3406)
    /gene="b10A"
    /note="b0774"
    complement(2117..3406)
    /gene="b10A"
    /EC_number="2.6.1.62"
    /function="enzyme; Biosynthesis of cofactors, carriers:
    Biotin"
    /note="t429; 100 pct identical to B10A-ECOLI SW: P12995"
    /codon_start=1
    /transl_table=11
    /product="7,8-diaminopelargonic acid synthetase"
    /protein_id="AAC73861.1"
    /db_xref="GI:1786991"
    /translation="MTMDLAFDQRIHMPYTSMTSPFLPYVVSAGCELLISDCRR
    LVDSSSWMAIHGYNHNPOLNAAMKSOIDKAMSHVQRTFTHAETLCRLVAMTPQP
    LECYFLADSSSAVEVAMKVALQYWKQAGARQFRTFRNGYQDTEFGASVCDPNS
    GMSLWKGYLPENLFAPAQSRMDGEMDERMVFARMAHREHIAVITEPTVQAG
    GMRVYHPWELKRIKICDRGILLADEIATGFRGKIFACCHAEIAPILLTGKAI
    TGGTMTLSATLTREVAETISGEAGCFMGPEMCPPLACAANASIALTESGDMCO
    QVADIEVOLREOLAPARDAEMVAVRVLGIVGVEPTTHPVNMAALOKFEVQGVWIRP
    FGKLIYMPYIILPOOLRLTAANRAVDETFECO"
    complement(2193..2221)
    /gene="b10A"
    /note="factor Sigma70; predicted +1 start at 807260"
    3411..3450
    /note="central position to biob promoter: -20"
    /bound_moiety="Biob documented site"
    3411..3450
    /note="central position to predicted promoter: -20; genetic
    evidence for the site"
    /bound_moiety="Biob predicted site"
    complement(3411..3450)
    /note="central position to predicted promoter:50"

promoter
    /bound_moiety="Biob predicted site"
    complement(3411..3450)
    /note="central position to biob promoter:50"
    /bound_moiety="Biob documented site"
    3413..3441
    /note="factor Sigma70; promoter biob; documented +1 at
    808525"
    complement(3447..3473)
    /note="factor Sigma70; promoter biob; documented +1 at
    808515"
    3493..4533
    /gene="b10B"
    /note="b0775"
    3493..4533
    /gene="b10B"
    /EC_number="2.8.1.-"
    /function="enzyme; Biosynthesis of cofactors, carriers:
    Biotin"
    /note="t046; 99 pct identical to B10B-ECOLI SW: P12996"
    /codon_start=1
    /transl_table=11
    /product="biotin synthesis, sulfur insertion?"
    /protein_id="AAC73862.1"
    /db_xref="GI:1786992"
    /translation="MAHREPWTLISQVTELEKPLDLLEFAQOVHROHFDPROVST
    LLSIKTGACPEDECKYCPOSSRYKTGLEARLVEQVLEASARKAKAAGSTRGMAW
    KNHERDMPYLEQMOVGVKAMGLEACMTIGTISESOARLANAGLIDYVHNLDTSPEF
    YGNIITTRTYOERLDLTKERYRDAGTKVSGGIVGCEYKDRAGLILLOLANPTPES
    VPINMLVKYKGPRLADNDVDADPDRITAVARIMPTSYKLSKREDMNQTAMC
    PMGANSITVCGCKLLTPNPEEDKQLQIRKGLINPOQAVIAGNEQOORLEQALMT
    PVDERYNAAL"
    4530..5684
    /gene="b10F"
    /note="b0776"
    4530..5684
    /gene="b10F"
    /EC_number="2.3.1.47"
    /function="enzyme; Biosynthesis of cofactors, carriers:
    Biotin"
    /note="t0384; 100 pct identical to B10F-ECOLI SW: P12998"
    /codon_start=1
    /transl_table=11
    /product="8-oxo-7-oxononanoate synthase"
    /protein_id="AAC73863.1"
    /db_xref="GI:1786993"
    /translation="MSWQEKINAALDARRAADALRRRYVPAQAGRWLVADROYLNF
    SSNDYLGISHPOIIRAMOGCAEFGICSGSGHVSQSVHQALEELFARMIGYRA
    LFIISGFANQAVIAAMAKEDRIADRLSHASILEASISPSORRRAHNDYHILAR
    LASPCGGQWVTEGVFSMDCSDAPLAELIOQYTOOHNGWMLVDDAHMETGYIGDGRG
    SCWLDKVRKELLVTFGKRGVSGAAVLCSSIVADYLLQFARHLIYSTSMPPAOAOL

Query Match
Best Local Similarity 100.0%; Score 25; DB 1; Length 11022;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ggcagtcacgcgcgcacattcgat 25
|||||
Db 3576 GCAGTGCATCCGCCACATTTCAT 3600

RESULT 13
AE005258
LOCUS AE005258 13501 bp DNA
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
OF 135.
ACCESSION AE005258 AE005174
VERSION AE005258.1 GI:12513751
KEYWORDS
SOURCE
ORGANISM Escherichia coli O157:H7 EDL933.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 13501)

```

AUTHORS Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J.J., Kirkpatrick, H.A., Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grobbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamou, S.K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A., and Blattner, F.R.

TITLE Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7

JOURNAL Nature 409 (6819), 529-533 (2001)

MEDLINE 21074935

PUBMED 11206551

REFERENCE 2 (bases 1 to 13501)

AUTHORS Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J.J., Kirkpatrick, H.A., Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grobbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamou, S.K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A., and Blattner, F.R.

TITLE Direct Submission

JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

FEATURES

source 1. 13501

Location/Qualifiers

/organism="Escherichia coli O157:H7 EDL933"

/strain="EDL933"

/serotype="O157:H7"

/db_xref="taxon:155864"

/note="Enterohaemorrhagic"

/note="O-island #36: Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655: Cryptic prophage CP-933K: Includes one copy of the 13 bp direct repeat that flanks the prophage"

66. 665

gene /gene="lomk"

/note="20981"

66. 665

/gene="lomk"

/function="putative membrane: Other or unknown (Phage or Prophage Related)"

/note="Residues 1 to 199 of 199 are 72.86 pct identical to residues 1 to 199 of 199 from Genpept 118 : g1175327891gb|AA63231.1|AF151091_2 (AF151091) lom [prophage P-ElDA]"

/codon_start=1

/transl_table=1

/product="putative outer membrane protein of prophage CP-933K"

CP-933K

/protein_id="AA655137.1"

/db_xref="GI:12513752"

/translation="MRKVCALILSAICLAVSGVPAMASEHSTLSAGYLHASTDAPG SDPLINGINRYEFTDGLITFSFYANAEDOKTHYSDTHEDYVNNRFSVWAG PSYRNEMISAYAMGAVASRYSTPSGDYFRYDKNKRDHDLTGSDARYSNTSLAW GAVOVNPTESYAVDAVATEXSGSGDKRIDGFIYGVGYK"

719. 2041

gene /gene="20982"

719. 2041

/gene="20982"

/function="putative structure: Structural component (Phage or Prophage Related)"

/note="Residues 164 to 440 of 440 are 68.79 pct identical to residues 381 to 645 of 645 from Genpept 118 : g1145854361gb|AAD25464.1|AF125520_59 (AF125520) putative tail fiber protein [Bacteriophage 933W]"

/codon_start=1

/transl_table=1

/product="putative tail component of prophage CP-933K"

/protein_id="AA655138.1"

/db_xref="GI:12513753"

/translation="MGMAVOISGVILKDGAKPIQCTIQLKARNSTXYVNVTVASE NPDAGRYSDVEGVSVTLVIEGPPSHAGTISYEDSQPTLNDPLFGAMTEDAR PEALRPEQVVEFAARHAEAKKNAGCAETASIGASAKASANAADTSADASE SAROAMASAKSASEASSSSASPAOKRASELOSATDAELSKTAEASAGAAADA TSTERKESASQSAESQRIAEADAVNRTPIYVGPGRGPGAGCGPGDGCGE

gene /gene="20984"

2004. 2312

/gene="20984"

2004. 2312

/function="orf: Other or unknown (Phage or Prophage Related)"

/note="Residues 1 to 102 of 102 are 98.03 pct identical to residues 28 to 129 of 129 from Genpept 118 : g1145854371gb|AAD25465.1|AF125520_60 (AF125520) hypothetical protein [Bacteriophage 933W]"

/codon_start=1

/transl_table=1

/product="unknown protein encoded by prophage CP-933K"

/protein_id="AA655139.1"

/db_xref="GI:12513754"

/translation="MRKAVIQALENDNINILKLMOSLGGCKHDDCENGSLITQLR LGPADILSESDENGIIPEODRVITQVVILDAKDKIOCVVRPIQILRADSTWENIGMK"

2489. 3469

gene /gene="20985"

2489. 3469

/gene="20985"

/function="orf: Other or unknown (Phage or Prophage Related)"

/note="Residues 13 to 326 of 326 are 55.73 pct identical to residues 18 to 311 of 336 from Genpept 118 : g1169603671gb|AA633527.1 (AA633527) hypothetical protein predicted by Glimmer [Salmonella typhimurium LT2]"

/codon_start=1

/transl_table=1

/product="unknown protein encoded by prophage CP-933K"

/protein_id="AA655140.1"

/db_xref="GI:12513755"

/translation="MLSPITRTTHNSVNIYVSSPCQTVSPAGKEKELVDEKPTILF QMPEPNERKDEVPILVNTKQHPYLDNVTNARIRSDRRTGTFYVGDPSVNNKAFS KLERDPENVMIIYREDVDFSMYDKRLSDIYHIIICEDRLTEKRDYELINLKELEK EISKAODSLISMVAKKRNHAMPFEFRMLALIKAGEIFRCYVNRKNGIISFEGCITLVD MDMLTGLKGTIVPDCGISMHVDRNSVNIENSALIVNSNHPALLEGISFMSKVD AHPYVDLGGVKKYTFNFTPLNHNHRCDFIEFNHPI INNTSQYTCSSW"

3530. 4522

gene /gene="20986"

3530. 4522

/gene="20986"

/function="orf: Other or unknown (Phage or Prophage Related)"

/note="No significant matches"

/codon_start=1

/transl_table=1

/product="unknown protein encoded by prophage CP-933K"

/protein_id="AA655141.1"

/db_xref="GI:12513756"

/translation="MKRPSLOSNNFNSPAGYSAPIAPNRAENAYADVIDGRIPL SAADSNVYESVITAVHDSRSRLIDQITQSTCEIKETIYVQDGPRIPEGAGNDNEEY VHIGCTIKRNEYELNEESSYKIDIDQISTCEIKETIYVQDGPRIPEGAGNDNEEY VSFVAPPTDSTYEMPSWDEGLIHETIHVYTSPPSDSNIELEPTEIARVAPDEIG WSPDFKGYAEERERHILRLRNALNRQAAHRENEAFERLGTISDRTEASPDFT EYSAVSNIGYFIQHDHPEGLAINDODANOIOLYHGAPYIFFGVDKHNQ"

5350. 6231

gene /gene="20989"

5350. 6231

/gene="20989"

/function="orf: Other or unknown (Phage or Prophage Related)"

/note="No significant matches"

/codon_start=1

/transl_table=1

/product="unknown protein encoded by prophage CP-933K"

/protein_id="AA655142.1"

/db_xref="GI:12513757"

/translation="MLSPYSVILGCSNNSLTNRNLTPDNRLVSSVRDAVHSDNGAOY

```

KGNRTYRVATDNKFCVTRSHSGCTNLLRLGPKCEISKEIEWIMNSPVSAAM
ETTSQSNRPDI.PPVDYAPPELPSVDYNLSPGVNIGKGAAYVEDAETKVI.KM
FTTSQSEVTSVRCFNOYVAGASKEIYGNNGDI.IG.RMKINGESILNTISSI.PAQ
ABHAIYDMFDRLEOKGILFVDITETNVLVDRAKNEFNPIDISSVWDRSMESQ.IMG
SVHGKODLIISVLSK1"
6462..7160
/gene="20990"
6462..7160
/gene="20990"
/feature="orf: Other or unknown (phage or prophage
related)"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG5143.1"
/db_xref="GI:12513758"
/transl_table="MRPISLNLVLHQSRRSSMSDTEISLVKASSVQWIKNPOLIK
OSTDHNITVOQIEALDLSGTEGTVLNLAIESTSRKSETVVHILNRSRGWAHRO
IDAENHRSDFHCNLTNAVEPCGEGISYVDFHATYFHELLHVFHNLNCRKHS
SRAESOKYSPILLFEPARTVIGARSEEVLSNKPHHEIGMPRTSYPDLSAL.IHDMT
VSLGFQVRLRHLT"
complement(7668..8144)
/gene="ydhb"
/note="ydhb"
/note="20992"
complement(7668..8144)
/gene="ydhb"
/function="orf: Unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf: hypothetical protein"
/protein_id="AAG5144.1"
/db_xref="GI:12513759"
/transl_table="MKLISNDLRDGRGLPHRHNENGMDYDGNISPLIANDVDPATK
SFVATCYDPDAPGSGMHWVYVNL.PADTRVLPGFGSGVAMPDGLQTRDPGKIG
YDGAAPKPGETHRYIFTVHLDVLRIDVDEGASGAMVGVHFSLASASITMFS"
complement(8203..9492)
/gene="p10a"
/note="p10a"
/note="20993"
complement(8203..9492)
CDS
Query Match 100.0%; Score 25; DH 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 0.00037;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 gcagtgatcgccagcatcgaat 25
|||||
db 9662 GCAGTGCATCGCCAGCATTTTCGAT 9666

RESULT 14
AP002553
LOCUS AP002553 297816 bp DNA BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION AP002553 BA000007
VERSION AP002553.1 GI:13360211
KEYWORDS
SOURCE
Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (sites)
Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C. and Shinagawa,H.
Complete nucleotide sequence of the prophage VP2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak

JOURNAL
MEDLINE
REFERENCE
AUTHORS
Genes Genet. Syst. 74 (5), 227-239 (1999)
20198780
2 (sites)
Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
Comparative analysis of the whole set of RNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
3 (sites)
Yokoyama,K., Makino,K., Kubota,Y., Matanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
Shinagawa,H.
Complete nucleotide sequence of the prophage VP1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
J. Infect. Dis. 182 (1-2), 127-139 (2000)
20564182
4 (sites)
Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shinagawa,H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
5 (bases 1 to 297816)
Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
Hayashi,T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail:kensgen-info.osaka-u.ac.jp,
URL:http://www.gen-info.osaka-u.ac.jp/, Tel.81-6-6879-8365,
Fax:81-6-6879-2047)
genome project.
FEATURES
Location/Qualifiers
1..297816
1.297816
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
79..1245
/gene="Ecs0753"
79..1245
/gene="Ecs0753"
/note="Ecs0753"
/note="similar to SUCC-ECOLI g11786948 percent identity
100 in 388 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="succinyl-CoA synthetase beta subunit"
/protein_id="BAB34176.1"
/db_xref="GI:13360212"
/transl_table="MNLHEYQAKOLFARYGL.PAPVGYACTTPREAEASAKIGAPWV
VKCVHAGRGKAGGVKVNSEKEDIRAFEAEMKRLVYOTDANOGPVNOLIVEAT
DIARELYGAVDSSRRVYMASTEGVEIEKVAEEMPH.IHKVNLADPLTGPPYOG
RELAFKLGLEKLIWOOFKIPMGJATFLELDIALIENPLVITKQSDITCLDGKLSA
DGNALFRPDLREKRDSDOEPRRQAQAWELNVALDNGTCWVNGAGLAKMTGDIIV
KLHGEPANFLDVGGGATKERVTAETKTIISDDKVAVLNIVGIVRCDLADGIIIG
AAVEGVNVPVVRLENNMELGAKKLADSLNTIAKGLTDAQAQVVAAVEGK"
1245..2114
/gene="Ecs0754"
1245..2114
/gene="Ecs0754"
/note="Ecs0754"
/note="similar to SUCC-ECOLI g11786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1

```


membrane-spanning protein; hsp1 superfamily proteins are associated with protein translocation; putative"

```

/codon_start=1
/transl_table=11
/product="Hsp1"
/db_xref="GI:151268"
/translation="MKRYINPLNMVALSMRSELVGAFVIAIVEMITPLPTGLID
VIAVNICISLLIMLMLHLPPLAFSTFPALLLTMRRLASVSTRLLILNDAG
HIVEAGQFVGVGNLAVGLVFLITVNFVITKGSERVAVGAFITLDMPGKMS
IDSIDLRLITVFEARKRRAELNKSOLFAMDCAMKFNCGIDAIASLITVAINMIGCI
SIGVLOHNMAGDALOLYTLITGDLIAQLPALLISVTSQMIIIRVNTGACYEANI
GREIAEOLITSPKAMIIASVAMLGFAIPGMEYGFITAIICGAGGILLOLRKPKA
DEORTAAVAPEMKEDILRTSPSRQFVLOTHPHGDSQATIALVSEIKRRNRRLVYV
GITTPSFIIEHYDDIAPDFRFTVYDPMLEKATPTQSHAVAEAROLEGENDLPAI PCN
TDROEDOWMLPAEOSGELNPVSTPLITERNERALQSCAPQIGLOETKALISMLKS
RQPLAEQWQVLTITREFSAVLQRLASCVPRLAIRVIAETLIEHCHEPRTNVLTVY
VRIALSOIYHQVCGAELQWLVTPSEGLRLDGLRQTOETEPFALSNETSOMLVOD
LHIAFPVPAEQAVLLVAQDLRSPLRTILREFEYHVPPLSAEISNAKVVVMPFTDL
EDDEPLDNEHNA"
BASE COUNT      772 a 1093 c 1035 g 703 t
ORIGIN
Query Match      76.0%; Score 19; DB 1; Length 3603;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcaggtgcacgcgcagcat 19
      |||||
Db 1522 GCAGGTGCATGCCGACAT 1504

```

RESULT 16
LOCUS MAV250020 2923 bp DNA BCT 26-SEP-2000
DEFINITION Mycobacterium avium subsp. paratuberculosis insertion sequence
IS900, Locus 7.
ACCESSION AJ250020
VERSION AJ250020.1 GI:8919146
KEYWORDS 2599 gene; 2600 gene; insertion element; insertion sequence IS900;
p43 gene; transposase.
SOURCE Mycobacterium avium subsp. paratuberculosis.
ORGANISM Mycobacterium avium subsp. paratuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinomycetales; Corynebacteriaceae;
Mycobacterium; Mycobacterium avium complex (MNC).
REFERENCE 1 (bases 1 to 2923)
AUTHORS Bull T.J., Hermon-Taylor, J., Pavlik, I., El-Zaatari, F. and Pizard, M.
TITLE Characterization of IS900 loci in mycobacterium avium subsp.
JOURNAL paratuberculosis and development of multiplex PCR typing
MEDLINE Microbiology 146 (Pt 9), 2185-2197 (2000)
REFERENCE 2 (bases 1 to 2923)
AUTHORS Bull T.J.
TITLE Direct Submission
JOURNAL Submitted (29-NOV-1999) Bull T.J., Surgery, St. George's Hospital
MEDLINE Medical School, London, SW17 0RE, UNITED KINGDOM
FEATURES
source 1..2923
/organism="Mycobacterium avium subsp. paratuberculosis"
/sub_species="paratuberculosis"
/db_xref="taxon:1770"
/complement(11..466)
/gene="2600"
complement(11..466)
/gene="2600"
/function="unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB96064.1"
/db_xref="GI:8919147"

translation="MYLAVETGVNDLPVLKGVAVATIEAVGMAVILVGFVADLLT
PCKRLQVFDPRRPNVAVGAMVITLVYITTAIANSISQAGVAVGVLGKSVI
LIGVALLAMHLLIPGSGHEHVEEQIHPGSAFVALLVAVGVTAAVS"
473..1922
/organism="Mycobacterium avium subsp. paratuberculosis"
/insertion_seq="IS900"
/sub_species="paratuberculosis"
/db_xref="taxon:1770"
686..1906
/gene="p43"
686..1906
/gene="p43"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="CAB96065.1"
/db_xref="GI:8919148"
/translation="MTVEVVAQPYWAGVDGKADHYCMYINDAQRLLSORVAND
ALLELIAVATTLADGEVYMAIDLNAGGALLIALLIAGORLYIPGVTHHAGS
YRGEKTDADKDAI IADQARMHDIQPLRAGDIIVEILITSRSDDIYADRTAIEP
NARPAGITLSALERAFVYKSRRAILITGYCPALRSAGCARVAAPLRKARNAD
TVATALQAMNQHSTVPGQOLAAYVVARLAEVVALDTEIGDTAMIEERRRHA
EILISMPGCVILGAEFLAATGDMAAAFASARLAVAGLAVPDSGRISGLKPR
RYDRLLRACVYSALVSTIRTPSSRTTYDRKRTKRRHQAVALARRRLVLMMLR
DRAVYHPATTTAA"
complement(1963..2394)
/gene="2599"
complement(1963..2394)
/gene="2599"
/function="unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB96066.1"
/db_xref="GI:8919149"
/translation="VSRNRLFYIAGALAAVAVCYVTGIILOKRIASVYAGHIEYA
RVNGTRRCYCGSSQVADPTLSEYSEAPARADDKTEYLRISNDIYIVGPNPCSI
RVPLSAGISHAEPFLICGFFPSPSGAGSTPGGPGTK"
BASE COUNT 473 a 1005 c 1000 g 445 t
ORIGIN

Query Match 72.0%; Score 18; DB 1; Length 2923;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcaggtgcacgcgcagca 18
 |||||
Db 127 GCAGGTGCATGCCGACCA 144

RESULT 17
LOCUS AF250776 5526 bp DNA BCT 31-JAN-2001
DEFINITION Uncultured bacterium pcosHE2 hypothetical 17.1 kDa protein in
modc-bioA intergenic region, DAPA-mannitolrasiferase BioA (bioA),
biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and
biotin biosynthesis protein BioC (bioC) genes, complete cds; and
dethiolobiotin synthetase BioD (bioD) gene, partial cds.
ACCESSION AF250776
VERSION AF250776.1 GI:12620124
KEYWORDS
SOURCE uncultured bacterium pcosHE2.
ORGANISM Bacteria; environmental samples.
REFERENCE 1 (bases 1 to 5526)
AUTHORS Entcheva, P., Liebl, W., Johann, A., Hartisch, T. and Streif, W.R.
TITLE Direct cloning from enrichment cultures and a reliable strategy for
JOURNAL isolation of complete operons and genes from microbial consortia for
MEDLINE Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
PUBMED 20575196
1133432
REFERENCE 2 (bases 1 to 5526)

AUTHORS Entcheva, P., Liebl, W. and Streitz, W.R.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
 Goettingen, Grisebachstr. 8, Goettingen 37077, Germany

FEATURES
 source Location/Qualifiers
 1..5526

/organism="uncultured bacterium pCosHE2"
 /db_xref="taxon:143797"
 /clone="pCosHE2"
 /note="unknown organism, cosmid clone derived from
 environmental consortium"
 complement(52..528)
 /note="ORF1"

CDS

/codon_start=1
 /transl_table=11
 /product="hypothetical 17.1 kDa protein in modC-bioA
 intergenic region"
 /protein_id="AAG60577.1"
 /db_xref="GI:12620125"
 /translation="MKLISNDLRDGDRLPHRHVNGMGYDGNISPLAMDVPAGTK
 SFVYTCIDPDAPFGSGMMHWYVNLDPDTRVLPDGRSGGLVAMPDGLQRTDGRKAG
 YDGAAPRGETHRYITVHALDIERIDVDSAGAMVGFNVHFLASASITAMFS"
 complement(587..1876)
 /gene="bioA"
 complement(587..1876)
 /gene="bioA"
 /note="7,8-diaminopelargonic acid
 synthetase-aminotransferase"
 /codon_start=1
 /transl_table=11
 /product="DAPA-aminotransferase BioA"
 /protein_id="AAG60578.1"
 /db_xref="GI:12620126"

CDS

/translation="MTTDLAFDORHIMHPYTSMTSPVPYVVSABCECELLSDGR
 LVDGSSWMAIHGYNHPOLNAAKMSQIDAASHVFGITAPAEILCRKLVAMPPOP
 LKCVPLADSGVAVAMKMALOYWQAKGEORFLPTRNHGDITFGAMVCDPDS
 MHSIKGLYPENLFAPAPQSRMDGEMDERDVGFLPMAARHETIAAIVIPYVAG
 GMRVHPMIRKICDREGILLADELAFGEFTGKFLACENHLEIPDLICGKAL
 TGGTWTLSATLTTRTREVATISNGEACGPMHPTGPNLACAAANASIALLESQMOO
 OVADTEVOLREOLAPARDAENYADRYVCAIGVETTPVNMALQKRFVQGWIRP
 FGLTYLMPYITLPOLQRLTAANVAVODETFECQ"
 1963..3003
 /gene="bioB"
 1963..3003
 /gene="bioB"
 /codon_start=1
 /transl_table=11
 /product="biotin synthase BioB"
 /protein_id="AAG60579.1"
 /db_xref="GI:12620127"
 /translation="MAHRPRMTLSOVTELEFEKPLDLLFEAOVHROHEFDPROVOST
 LLSITGACIL...KYCROSSRYKGLFAERLMEVQVLESARKAKAAGSTRPCMAAM
 KPNEDMPTLROMVOGVKDLGLACMTIGTISPOAORLANAGIDYNNHNLTPSPF
 YGNLITTRTYOERDLTLEKVRDAGIKVSGIVGEGYKDRAGILLDLANLPTPES
 VPINMLVKVGTPLADNDVDAFDLRTIARVARIMPTISYVRLSGREDMNQTOAMC
 PMAGANSIFYGCKLTTTPNEEDKDLQLFRKGLNPOOTAVLAGNDOOQRLLEALMT
 PDDEYNAAL"
 3000..4154
 /gene="bioF"
 3000..4154
 /gene="bioF"
 /note="8-amino-7-oxononanoate synthase"
 /codon_start=1
 /transl_table=11
 /product="KAPA synthetase BioF"
 /protein_id="AAG60580.1"
 /db_xref="GI:12620128"

gene

CDS

/translation="MSWQKINNALDARAADALRRRYVAGAGHVLVADROVILNF
 SSNDYGLSHHPQIRAMOGAEOFGISGSGSHVSGVSHALAEELAEMLCYRA
 LFIISGPAANOAVITAMMAKEDRIADRLSHASLLEASISPSOLRFAHNDVTIHLR
 LMSPCGOOLVYTEGVFSMDSDAPLAEIOQVTOOHGMWMDAHDAGHGVIGEGRG
 SCMLQKRPPELLVYTGKGGVSGCAVLCSSYADYLLQFAHLLYSTSPMPAOAOL
 RASLAVIRDEGAREKRLVSLIAHFRAGVDLPTFLADSCSAIOPILYDGNRSALOL

gene

CDS

gene 4141..4896
 CDS 4141..4896
 /gene="bioC"
 /note="reaction step prior to synthesis of pimeloyl-CoA"

/codon_start=1
 /transl_table=11
 /product="biotin biosynthesis protein BioC"
 /protein_id="AAG60581.1"
 /db_xref="GI:12620129"
 /translation="MATVVKQALIAAFGRALYEQHDLQROSDAVLLAMPKRYT
 HVLDAAGCGMMTRHRRERHAQVTLADSPMLVQARKDADHDLADIELPLATA
 TEDLAWNSLAVOMCNSLSTALRELYRVVSKGVAFYTLVQSGPELHQAOWADERP
 HANRFLPDEIEQSLNGVYOHNIQPTILMPDALSAMRSKIGATHLDEGRDRL
 TRSOLQRLQAMPOOGRPLTYHLFLGVARE"
 4883..5526
 /gene="bioD"
 4883..5526
 /gene="bioD"
 /note="DTP synthetase"
 /codon_start=1
 /transl_table=11
 /product="dethiobiotin synthetase BioD"
 /protein_id="AAG60582.1"
 /db_xref="GI:12620130"
 /translation="MLVSKRYFVYGTDEYKTVASCALLQAKAGYFTAGKPYVAS
 GSEKTEGLNSDALALORNSLQLDYATVNPYTFAPETSPHLSAEOGRPIESSYMS
 SGIRALEQOQADWLVEGAGGMPPLSDTFEAMVYVQEDLPVLYVGLGGINHML
 TAOIQRAGITLQGWVANDVTPPKRHAETITLTMTAIRAPILGELPMLAENPE"

CDS

BASE COUNT 1274 a 1507 c 1567 g 1178 t
 ORIGIN

Query Match 72.0% Score 18; DB 1; Length 5526;
 Best Local Similarity 100.0%; Pred. No. 4;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 8 catcgccagcatlccgat 25
 Db 2053 CATGCCAGCATTTCCGAT 2070
 |||||

RESULT 18
 G01446 479 bp DNA STS 19-AUG-1999
 LOCUS DMO466 drosophila pl library Drosophila melanogaster STS genomic
 DEFINITION clone DS07967 Sp6, sequence tagged site.
 G01446
 ACCESSION G01446.1 GI:684849
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 fruit fly.
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 RUBIN, G.
 Berkeley Drosophila Genome Project
 Unpublished (1994)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Berkeley Drosophila Genome Project
 Primer A: CTCCTACTGGGACTTGCCTA
 Primer B: CACCAAGCCATATCCCA
 STS size: 192
 PCR Profile:
 Annealing: 58 degrees C PCR Cycles: 30
 Protocol:
 Template: P1 Library Pools
 Primer: 1 uM each
 dNTPs: 250 uM each
 Tag Poly: 0.05 units/ul

Total Vol: 15 uL

Buffer:

MgCl₂: 1.5mM
KCl: 50 mM
Tris-HCl: 50 mM
pH: 8.3
Gelatin: .001 %

The pl library has been distributed to 16 regional sites. A list of these sites is available from Flybase, via anonymous ftp to ftp.bio.indiana.edu in the file flybase/allied.data/genome-projects/1bl/LBL.doc.

FEATURES

source
1. 479
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="DS07967"
/note="Vector: PAd10Sact1: The pl library was made by D. Smoller in D. Hartl's lab (see Smoller et al., Chromosoma 100: 487). Chromosomal position was mapped by polytene chromosome in situ hybridization in the Hartl lab (see Hartl et al., PNAS 91: 6824). STS's were generated by sequencing the ends of the Drosophila insert in these pl clones, by the W. Kimmerly, C. Martin, and M. Palazzolo lab at LBL."
STS
primer_bind complement(233..250)
BASE COUNT 113 a 120 c 97 g 145 t 4 others
ORIGIN

Query Match 68.0%; Score 17; DB 11; Length 479;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gcatgccagcatctcg 23
Db 396 GCATGCCAGCATTTGG 412
|||||

RESULT 19
AE007981/c LOCUS 10464 bp DNA BCT 14-AUG-2001
DEFINITION Agrobacterium tumefaciens strain C58 circular chromosome, section 39 of 254 of the complete sequence.
ACCESSION AE007981 AE007869
VERSION AE007981.1 GI:15155357
KEYWORDS
SOURCE Agrobacterium tumefaciens.
ORGANISM Agrobacterium tumefaciens; alpha subdivision; Rhizobiaceae group; Bacteria; Proteobacteria; Rhizobium.
1 (bases 1 to 10464)
Hinkle,G., Slater,S.C. and Goodner,B.
Complete Genome Sequence of Agrobacterium tumefaciens C58 (Rhizobium radiobacter C58), the Causative Agent of Crown Gall Disease in Plants
Unpublished
2 (bases 1 to 10464)
Hinkle,G., Slater,S.C. and Goodner,B.
Direct Submission
Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street, Cambridge, MA 02139, USA
Location/Qualifiers
1. 10464

FEATURES
source
/organism="Agrobacterium tumefaciens"
/strain="C58"
/db_xref="taxon:358"
complement(162..1067)
/gene="AGR_C_768"
complement(162..1067)
CDS
gene
complement(162..1067)

/gene="AGR_C_768"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_768p"
/protein_id="AAK86250.1"
/db_xref="GI:15155358"
/translation="MDLTOTAAVEIAKLALTNNAIVQVAPADAKGIPSTYPALDPTSGRLADVAIFAPMRTREPKEKQAYAEETLESFVALVERHOTENSATPATWMDPTTAVIDYHGGDDPDNGKHRYVAFPLESEKAKAQSTNGKALTONERAFETIEHIAELASPSDEYKDELELFKAVATPNEVLVSQGLQINAEIRVKTAVKLOTGSQIVFEEDHKMANGDPTTGVGFVLNIAIPFQGESIRLPLVRLRYLRREGKLSMTCLYRPPHTITKAVTALHTAELIDLPKAGKPEMSA"
complement(1127..1459)
/gene="AGR_C_769"
/gene="AGR_C_769"
complement(1127..1459)
/gene="AGR_C_769"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_769p"
/protein_id="AAK86251.1"
/db_xref="GI:15155359"
/translation="MKKIDASTITGLENGQINPAFSAEIGQTLLEKSDSEANPMQSFSGSVTLKLSLVKQGMVTIADMSKTPKLPKNSVWVEDALSTEHPRDHMFIPREVSAS"
complement(1490..1702)
/gene="AGR_C_771"
complement(1490..1702)
/gene="AGR_C_771"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_771p"
/protein_id="AAK86252.1"
/db_xref="GI:15155360"
/translation="MDTGVKTRORALYLRHCAEYAEENFVPTTDEDEDETPKPLHARLREANAEATAASALEMNDLMQYADELNRATMDAKKLDMDAFTGFCGMAA"
complement(1750..2061)
/gene="AGR_C_774"
complement(1750..2061)
/gene="AGR_C_774"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_774p"
/protein_id="AAK86253.1"
/db_xref="GI:15155361"
/translation="MDTGVKTRORALYLRHCAEYAEENFVPTTDEDEDETPKPLHARLREANAEATAASALEMNDLMQYADELNRATMDAKKLDMDAFTGFCGMAA"
complement(2117..2350)
/gene="AGR_C_776"
complement(2117..2350)
/gene="AGR_C_776"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_776p"
/protein_id="AAK86254.1"
/db_xref="GI:15155362"
/translation="MRKARPTQSGDGTGMRINTVNDDELTOAMADATQVGESECTKADLRFWTADIEIRCGDAIATARHMDRVQDARAA"
complement(2515..2928)
/gene="AGR_C_779"
complement(2515..2928)
/gene="AGR_C_779"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_779p"
/protein_id="AAK86255.1"
/db_xref="GI:15155363"


```

misc_feature /note="assembly_fragment"
30589..32514
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
32615..36175
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
36276..38734
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
38835..40927
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
41028..43453
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
43554..46744
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
46845..50369
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
50470..56799
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
56900..57617
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
57718..68517
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
68618..82793
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
82894..96371
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
96472..112208
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
112309..127004
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
127105..147474
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
147575..174126
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
174227..183937
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
clone_end:T7
vector_side:right"

```

```

BASE COUNT 47977 a 43626 c 42743 g 47258 t 2333 others
ORIGIN

```

```

Query Match 68.0%: Score 17; DB 2; Length 183937;
Best Local Similarity 100.0%: Pred. No. 14;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 6 tgcacgcacgacatttc 22
|||||
Db 90965 TGCATCCGACGACATTTC 90981

```

```

RESULT 22
AC092401 185200 bp DNA INV 04-JUL-2001
LOCUS Drosophila melanogaster, chromosome X, region 13E-13F, BAC clone
DEFINITION BACR25C18, complete sequence.
ACCESSION AC092401
VERSION AC092401.1 GI:14595766
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 185200)
Celisner,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frisze,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalili,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
2 (bases 1 to 185200)
Celisner,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frisze,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalili,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome X, region 13E-13F
Unpublished

```

```

TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (04-JUL-2001) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US
Sequence submitted by:
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu.

```

```

FEATURES
source
Location/Qualifiers
1..185200
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/map="13E-13F"
/clone="BACR25C18 (D1403)"
/clone_lib="RPCT-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBAC3.6)"

```

```

BASE COUNT 52552 a 40775 c 40327 g 51546 t
ORIGIN

```

```

Query Match 68.0%: Score 17; DB 3; Length 185200;
Best Local Similarity 100.0%: Pred. No. 14;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 7 gcatcgccacgacatttc 23
|||||
Db 88505 GCATCCGACGACATTTCG 88521

```

```

RESULT 23
AC013431 191558 bp DNA INV 31-MAY-2001
LOCUS Drosophila melanogaster, chromosome X, region 13E-13F, BAC clone
DEFINITION BACR22H11, complete sequence.
ACCESSION AC013431
VERSION AC013431.9 GI:14269669
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 191558)
Celisner,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,

```

Holt, R.A., Evans, C.A., Gocayne, J.D., Amanalides, P.G., Brandon, R.C., Rogers, J., An, H., Baldwin, D., Bonzon, J., Beeson, K.Y., Busam, D.A., Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, V., Dou, L.E., Doyle, C., Dresnek, D., Farfan, D., Gonzalez, S., Fris, E., Galle, R.F., Garg, N.S., George, R.A., Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattel, B., McIntosh, T.C., Moynihan, V., Murphy, B., Nelson, C., Nelson, K.A., Nunn, J., Paciel, J., Paragay, V., Park, S., Patel, S., Pfeiffer, B., Phoonanavong, S., Plittman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Swirskas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M., and Venter, J.C. 1995. Sequencing of *Drosophila* chromosome X, region 13E-13F. Unpublished

TITLE
JOURNAL
AUTHORS
REFERENCE
2 (bases 1 to 191558)
Celisner, S.E., Abdeyani, A., Arcata, T.T., Baxter, E., Blazej, R.G., Butenheff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hinkle, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Katta, K., Kearney, L., Lee, B., Lewis, S., Li, P., Ling, H., Moshrefi, A.R., Moshrefi, M., Nixon, K., Paciel, J.M., Park, S., Pfeiffer, B., Richards, S., Sethi, H., Swirskas, R.R., Wan, K.H., Webster, D., Woolley, P., Yang, S., Yee, M., Yu, C., and Rubin, G.M.
Direct Submission
Submitted (11-NOV-1999) *Drosophila* Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
On May 31, 2001 this sequence version replaced gi:6838815.
Sequence submitted by:
Berkeley *Drosophila* Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.berkeley.edu.
Location/Qualifiers
1..191558

FEATURES
source
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="X"
/map="13E-13F"
/clone="BACR22H11 (D1191)"
/clone_id="RPC1-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial ECORI in
PBAC3.6)"
BASE COUNT
54283 a 41835 c 41651 g 53789 t
ORIGIN

Query Match 68.0%; Score 17; DB 3; Length 191558;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 7 gcacgcacgattcg 23
|||||
Db 67235 GCATGCCAGCATTCG 67251

RESULT 24
AE003500
LOCUS
DEFINITION
AE003500 *Drosophila melanogaster* genomic scaffold 14200001386053 section 17
of 30, complete sequence.
ACCESSION
AE003500.2 GI:10728273
VERSION
KEYWORDS
HTG.
SOURCE
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

REFERENCE

AUTHORS

Muscomorpha: Ephydroidea: Drosophilidae: *Drosophila*.

1 (bases 1 to 327446)
Adams, M.D., Celisner, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanalides, P.G., Scher, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.C., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Wan, K.H., Doyle, C., Baxter, E.G., Heit, G., Nelson, C.R., Gabor, Miklos, G.L., Adell, J.F., Abdeyani, A., An, H.J., Andrews, P., Pflanz, C., Baldwin, D., Ball, W.R., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Bertram, B.P., Bhandari, D., Bolshakov, S., Borkov, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottler, P., Butts, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Meyers, A.D., Dew, I., Dietz, S.M., Dodson, K., Dou, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Fertler, S., Fleischmann, W., Foster, C., Gabriellian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, E., Karpen, G.H., Ke, Z., Kennis, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Matei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Mishina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Paciel, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J.D., Puri, V., Reese, M.G., Reiner, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Sheng, B.C., Siden-Kimios, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Swirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wasserman, D.A., Weinstock, G.M., Weissbach, J., Williams, S.M., Woodage, T., Zaveri, J.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibson, R.A., Myers, E.W., Rubin, G.M., and Venter, J.C.
The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)
2 (bases 1 to 327446)
Adams, M.D., Celisner, S.E., Gibbs, R.A., Rubin, G.M., and Venter, J.C.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7293108.
COMMENT
FEATURES
source
Location/Qualifiers
1..327446
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="X"
complement(join(772..903,968..1504,1572..2077,2153..2821,4145..>4380))
/gene="CG9281"
/product="CT26402"
/db_xref="FLYBASE:FBan0009281"
/db_xref="FLYBASE:FBgn0030672"
complement(join(772..903,968..1504,1572..2077,2153..2821))
/gene="CG9281"
/note="CG9281 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0009281"
/db_xref="FLYBASE:FBgn0030672"
/protein_id="AAFA8493.1"
/db_xref="GI:7293109"

MEDLINE 99392457
 REFERENCE 2 (bases 1 to 1186)
 AUTHORS Thomas,J.D.
 TITLE Direct Submission
 JOURNAL Submitted (29-APR-1999) Biological Sciences, University of Warwick,
 Gibbet Hill Road, Coventry CV4 7AL, UK

FEATURES
 source Location/Qualifiers
 1..1186
 /organism="Pectobacterium carotovorum subsp. carotovorum"
 /strain="SCRI193"
 /sub_species="carotovora"
 /db_xref="taxon:555"
 <1..189
 /gene="xerD"
 <1..189
 /gene="xerD"
 /note="similar to Escherichia coli XerD
 integrase/recombinase in Swissprot Accession Number
 P21891"
 /codon_start=1
 /transl_table=11
 /evidence="not_experimental"
 /product="xerD"
 /protein_id="AAD47611.1"
 /db_xref="GI:5712697"
 /translation="KLSPHVLRHAFATHLNHCAGFTGRITDALGAQRSPFDADLHKR
 DGAETRSAAHVRHRT"
 222..1031
 /gene="dsbc"
 222..227
 /gene="dsbc"
 246..251
 /gene="dsbc"
 258..264
 /gene="dsbc"
 /note="putative"
 315..377
 /gene="dsbc"
 /note="putative sec-dependent export signal"
 315..1031
 /gene="dsbc"
 /function="periplasmic disulfide bond formation"
 /note="disulfide isomerase; similar to Erwinia
 chrysanthemi and Escherichia coli DsbC proteins in
 Swissprot Accession Numbers P39691 and P21892"
 /codon_start=1
 /transl_table=11
 /product="dsbc precursor"
 /protein_id="AAD47610.1"
 /db_xref="GI:5712696"
 /translation="MKKGLLSLVATNPNFAHADDAAIKOTLHAGYARGGDPAFA
 DGMKVTIDSGVLTISEGQAFAPALRERTMPVNTNKLIGKLDALQEMIVY
 KAQERHATVTEFTISICGYCHKLHEQMKDYNALGITVRYLAFPGCMKSPAKDMQSI
 WCVAENKAFDSAMKGDATSAATCKTIDIAHVOGLGIFGVGTPAIVIQDGLVVGQY
 GPEMLAMEAHNSAKRTGG"
 378..1028
 /gene="dsbc"
 /product="dsbc"
 659..680
 /gene="dsbc"
 /note="Region: putative CPHC motif typical of the active
 sites of disulfide isomerases"
 1044..>1186
 /gene="recJ"
 1044..>1186
 /gene="recJ"
 /note="theoretical protein; similar to Erwinia
 chrysanthemi and Escherichia coli RecJ single-stranded
 DNA-specific exonucleases in Swissprot Accession Numbers
 P39693 and P21893"
 /codon_start=1
 /transl_table=11
 /evidence="not_experimental"

BASE COUNT 295 a 315 c 323 g 253 t
 ORIGIN

Query Match 64.0%; Score 16; DB 1; Length 1186;
 Best local Similarity 100.0%; Pred. No. 58;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gcatgcacgacatc 22
 |||||||||||||
 Db 993 GCATGCCACGATTTTC 978

RESULT 26
 AK024397/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

AK024397 4674 bp mRNA PRI 29-SEP-2000
 Homo sapiens cDNA FLJ14335 fis, clone PLACE4000411, highly similar
 to Homo sapiens mRNA: cDNA DKFZP586D0624 (from clone
 DKFZP586D0624).
 AK024397
 AK024397.1 GI:10436775
 oligo capping; fis (full insert sequence).
 Homo sapiens placenta cDNA to mRNA, clone_11b: PLACE4
 clone: PLACE4000411.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (sites)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 COMMENT

Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
 Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
 Wagaatsuma,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,
 Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
 Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
 Ishii,S., Kawai,Y., Saito,K., Yamamoto,D., Wakamatsu,A.,
 Nakamura,Y., Nagahara,K., Masuko,Y., Ninomiya,K. and Iwayanagi,T.
 NEDO human cDNA sequencing project
 Unpublished (2000)
 2 (bases 1 to 4674)
 Isogai,T. and Otsuki,T.
 Direct Submission
 Submitted (23-AUG-2000) to the DDBJ/EMBL/Genbank databases, Takao
 Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
 Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp,
 Tel:81-438-52-3951, Fax:81-438-52-3952)
 NEDO human cDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan; cDNA full insert
 sequencing; Research Association for Biotechnology; cDNA library
 construction. 5'- & 3'-end one pass sequencing and clone selection;
 Helix Research Institute (supported by Japan Key Technology Center
 etc.) and Department of Virology, Institute of Medical Science,
 University of Tokyo.

FEATURES
 source Location/Qualifiers
 1..4674
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="PLACE4000411"
 /clone_11b="PLACE4"
 /tissue_type="placenta"
 /note="cloning vector: pME18SFL3"
 608..1273
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAB14908.1"
 /db_xref="GI:10436776"
 /translation="MLPSLYKLGMTLVBPQKRELSAAAIKRIATATAKSGKLLQVY
 LKSPKRIITDNLTNQLTENSTVRICTADKMDKVFAYIAQSHNQSLTECHAFI
 CTRKMAQAVTLTVAQAFKVAEFEEFGVSEEEKERKRAQEGDVYGAQDCPTPLKS
 LVATGNLLDEETAKAPLSTVSANTNTMDEVPRPQALSSGVVWVSGCAVSSVILSL

BASE COUNT 991 a 1342 c 1239 g 1102 t
ORIGIN

Query Match 64.0%; Score 16; DB 9; Length 4674;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

4 ggtgcatccacacat 19
|||||
Db 2446 GGTGCATCCACACAT 2431

RESULT 27 AF206717 11910 bp DNA SYN 12-SEP-2000
LOCUS AF206717/c Shuttle vector p13 hypothetical proteins, Htru (htru), Ardu (ardu),
DEFINITION Repl (repu), and Resu (resu) genes, complete cds.
ACCESSION AF206717 GI:7262998
VERSION AF206717.1
KEYWORDS
SOURCE Shuttle vector p13.
ORGANISM Shuttle vector p13.
REFERENCE 1 (bases 1 to 11910)
AUTHORS Melina, R. and Lidstrom, M.E.
TITLE Characterization of the minimal replicon of a cryptic Deinococcus
JOURNAL radiodurans SARK plasmid and development of versatile Escherichia
coli-D. radiodurans shuttle vectors
APPL. Environ. Microbiol. 66 (9), 3856-3867 (2000)
MEDLINE 20422197
PUBMED 10966401
2 (bases 1 to 11910)
REFERENCE
AUTHORS Melina, R. and Lidstrom, M.E.
TITLE Direct Submission
JOURNAL Submitted (18-NOV-1999) Chemical Engineering, University of
Washington, Benson Hall, PO Box 351750, Seattle, WA 98195, USA
FEATURES
SOURCE location/Qualifiers
1.11910
/organism="Shuttle vector p13"
/specific_host="Deinococcus radiodurans SARK"
/db_xref="taxon:115340"
/note="derived from pUE10"
105..593
/note="orfA"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4052.1"
/db_xref="GI:7263004"
/translation="MGCOLVEGGDGVPPNDGGQODVHRLPLPGVPRKDFAMLRG
/translating="MGCOLVEGGDGVPPNDGGQODVHRLPLPGVPRKDFAMLRG
GKERONDGPDISMPKRRVTEBAHPDPDGGGVASGEARGFSPDPVLAPEKGEDEH
EOLDVLAELREVRCGEAAGOLGSDIGRRKVFVSVKRTSSPACNCRVILMLMOMQRPK
AV"
complement(257..683)
/note="5 mismatches with sequence deposited in GenBank
Accession Number M94966"
649..653
659..1078
/note="orfB"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4047.1"
/db_xref="GI:7262999"
/translation="MREVAFDSEPSQALKAITLGRKAAGLEVMYIEPPVSG
FVOMAGLNDVREMLRQNAQAOQTAAQTLLIGAEVQGRVOTLGRPIEVLTEAAGV
DLVYGTGGRGMDRIILGSVETTLRRAEPLLVVR"
1187..2436
/gene="htru"

RBS
CDS

1187..1192
/gene="htru"
1198..2436
/gene="htru"
/transl_table=1
/codon_start=1
/product="htru"
/protein_id="AAFA4048.1"
/db_xref="GI:7263000"
/translation="MKTAMPILTLALLAGCTDPESTRQTTSSSRKATTTMPARGA
ARDSSSPGAGTASPDQRLIEDNTIDVMDQGVVYTRDQSGGALFESBP
LAPPAEDQGPFGSGGFLIDREGHILTNVHIDADIRILQNRQDPAVVG
TAPVYDALLRARGARRGAGGARGSPAGRKDCGAPGLEFTYQGLITFAVK
RVIPMGVESLPONSQVTDAAINPCNSGCPVANSRGVGNQIISPAVAGVQONA
VGGAIPVNVKSLPLRAGEETVPRIGTIVRNQALPPSARALGIPRQGVLOS
VEGTPAAGLARGPRSORPPGDIRLGGDITTAVDGGEVSTVODLQVLLGKQSD
EVLILTRDQCTLQRLTLLS"
3113..3118
3127..3678
/note="orfC, proline-rich protein"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4053.1"
/db_xref="GI:7263005"
/translation="MTFNCAAMASRSTRSGRCATRSRSLGVGSESAVPAATV
ARLPVPPVRSAPFLGHDSDMHRPAPSGAARLPVPPVRSVGVGCSAPVLSRLP
RSTPHIRKISFEWRCFSASPLLRGQIVPGRAGGPPRLRPHWRPARACFP
AARPLVSLSCICAVTLVLLSCI"
3691..4038
/note="orfD"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4054.1"
/db_xref="GI:7263006"
/translation="MTTAPLPSASAVGSRGSAVAVSGPGAAWYMCNRRTAAR
ATLRIPASAPSTTAATVAVAMVAGRLRQVRRACGVLMKCSPPRAQIAAAG
WMAVIVATGRRQ"
4090..4551
/note="orfE"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4055.1"
/db_xref="GI:7263007"
/translation="MAAPTRKHSPPPWLLPLAGLMSVAPVALTTPPLPPPCSAI
TTPPPACCPVLTVPRTSPVGRSSTPPSPCFPPRGSAPARBMPCGWSRAA
CPFSLAPLRPPRPARPGRPPQAPFAGSGRRRPPACSSFPAPARLSFI"
4666..4670
4678..5364
/note="orfF"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4056.1"
/db_xref="GI:7263008"
/translation="MSELTFRUSPHARALVALGADHTAPAPARAADIYMOQISGP
APVTDAPRSAPARTLRPLRGVYGRWDSAPLPTARACKLNLTLTAGADPITMAR
ALTRGDTAAEQITCEGMQLDIQNRRAQLAAQOQOQAPAPAPLTLRTQARLHK
QIAGIWTDERVTOAKIARRPVESLITALLNODBAALLVVAABEARDRAAVLAQOD
PAAGCALIR"
5407..5973
/gene="ardu"
5407..5412
/gene="ardu"
5422..5973
/gene="ardu"
/codon_start=1

CDS

1187..1192
/gene="htru"
1198..2436
/gene="htru"
/transl_table=1
/codon_start=1
/product="htru"
/protein_id="AAFA4048.1"
/db_xref="GI:7263000"
/translation="MKTAMPILTLALLAGCTDPESTRQTTSSSRKATTTMPARGA
ARDSSSPGAGTASPDQRLIEDNTIDVMDQGVVYTRDQSGGALFESBP
LAPPAEDQGPFGSGGFLIDREGHILTNVHIDADIRILQNRQDPAVVG
TAPVYDALLRARGARRGAGGARGSPAGRKDCGAPGLEFTYQGLITFAVK
RVIPMGVESLPONSQVTDAAINPCNSGCPVANSRGVGNQIISPAVAGVQONA
VGGAIPVNVKSLPLRAGEETVPRIGTIVRNQALPPSARALGIPRQGVLOS
VEGTPAAGLARGPRSORPPGDIRLGGDITTAVDGGEVSTVODLQVLLGKQSD
EVLILTRDQCTLQRLTLLS"
3113..3118
3127..3678
/note="orfC, proline-rich protein"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4053.1"
/db_xref="GI:7263005"
/translation="MTFNCAAMASRSTRSGRCATRSRSLGVGSESAVPAATV
ARLPVPPVRSAPFLGHDSDMHRPAPSGAARLPVPPVRSVGVGCSAPVLSRLP
RSTPHIRKISFEWRCFSASPLLRGQIVPGRAGGPPRLRPHWRPARACFP
AARPLVSLSCICAVTLVLLSCI"
3691..4038
/note="orfD"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4054.1"
/db_xref="GI:7263006"
/translation="MTTAPLPSASAVGSRGSAVAVSGPGAAWYMCNRRTAAR
ATLRIPASAPSTTAATVAVAMVAGRLRQVRRACGVLMKCSPPRAQIAAAG
WMAVIVATGRRQ"
4090..4551
/note="orfE"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4055.1"
/db_xref="GI:7263007"
/translation="MAAPTRKHSPPPWLLPLAGLMSVAPVALTTPPLPPPCSAI
TTPPPACCPVLTVPRTSPVGRSSTPPSPCFPPRGSAPARBMPCGWSRAA
CPFSLAPLRPPRPARPGRPPQAPFAGSGRRRPPACSSFPAPARLSFI"
4666..4670
4678..5364
/note="orfF"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4056.1"
/db_xref="GI:7263008"
/translation="MSELTFRUSPHARALVALGADHTAPAPARAADIYMOQISGP
APVTDAPRSAPARTLRPLRGVYGRWDSAPLPTARACKLNLTLTAGADPITMAR
ALTRGDTAAEQITCEGMQLDIQNRRAQLAAQOQOQAPAPAPLTLRTQARLHK
QIAGIWTDERVTOAKIARRPVESLITALLNODBAALLVVAABEARDRAAVLAQOD
PAAGCALIR"
5407..5973
/gene="ardu"
5407..5412
/gene="ardu"
5422..5973
/gene="ardu"
/codon_start=1

CDS

1187..1192
/gene="htru"
1198..2436
/gene="htru"
/transl_table=1
/codon_start=1
/product="htru"
/protein_id="AAFA4048.1"
/db_xref="GI:7263000"
/translation="MKTAMPILTLALLAGCTDPESTRQTTSSSRKATTTMPARGA
ARDSSSPGAGTASPDQRLIEDNTIDVMDQGVVYTRDQSGGALFESBP
LAPPAEDQGPFGSGGFLIDREGHILTNVHIDADIRILQNRQDPAVVG
TAPVYDALLRARGARRGAGGARGSPAGRKDCGAPGLEFTYQGLITFAVK
RVIPMGVESLPONSQVTDAAINPCNSGCPVANSRGVGNQIISPAVAGVQONA
VGGAIPVNVKSLPLRAGEETVPRIGTIVRNQALPPSARALGIPRQGVLOS
VEGTPAAGLARGPRSORPPGDIRLGGDITTAVDGGEVSTVODLQVLLGKQSD
EVLILTRDQCTLQRLTLLS"
3113..3118
3127..3678
/note="orfC, proline-rich protein"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4053.1"
/db_xref="GI:7263005"
/translation="MTFNCAAMASRSTRSGRCATRSRSLGVGSESAVPAATV
ARLPVPPVRSAPFLGHDSDMHRPAPSGAARLPVPPVRSVGVGCSAPVLSRLP
RSTPHIRKISFEWRCFSASPLLRGQIVPGRAGGPPRLRPHWRPARACFP
AARPLVSLSCICAVTLVLLSCI"
3691..4038
/note="orfD"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4054.1"
/db_xref="GI:7263006"
/translation="MTTAPLPSASAVGSRGSAVAVSGPGAAWYMCNRRTAAR
ATLRIPASAPSTTAATVAVAMVAGRLRQVRRACGVLMKCSPPRAQIAAAG
WMAVIVATGRRQ"
4090..4551
/note="orfE"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4055.1"
/db_xref="GI:7263007"
/translation="MAAPTRKHSPPPWLLPLAGLMSVAPVALTTPPLPPPCSAI
TTPPPACCPVLTVPRTSPVGRSSTPPSPCFPPRGSAPARBMPCGWSRAA
CPFSLAPLRPPRPARPGRPPQAPFAGSGRRRPPACSSFPAPARLSFI"
4666..4670
4678..5364
/note="orfF"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4056.1"
/db_xref="GI:7263008"
/translation="MSELTFRUSPHARALVALGADHTAPAPARAADIYMOQISGP
APVTDAPRSAPARTLRPLRGVYGRWDSAPLPTARACKLNLTLTAGADPITMAR
ALTRGDTAAEQITCEGMQLDIQNRRAQLAAQOQOQAPAPAPLTLRTQARLHK
QIAGIWTDERVTOAKIARRPVESLITALLNODBAALLVVAABEARDRAAVLAQOD
PAAGCALIR"
5407..5973
/gene="ardu"
5407..5412
/gene="ardu"
5422..5973
/gene="ardu"
/codon_start=1

CDS

1187..1192
/gene="htru"
1198..2436
/gene="htru"
/transl_table=1
/codon_start=1
/product="htru"
/protein_id="AAFA4048.1"
/db_xref="GI:7263000"
/translation="MKTAMPILTLALLAGCTDPESTRQTTSSSRKATTTMPARGA
ARDSSSPGAGTASPDQRLIEDNTIDVMDQGVVYTRDQSGGALFESBP
LAPPAEDQGPFGSGGFLIDREGHILTNVHIDADIRILQNRQDPAVVG
TAPVYDALLRARGARRGAGGARGSPAGRKDCGAPGLEFTYQGLITFAVK
RVIPMGVESLPONSQVTDAAINPCNSGCPVANSRGVGNQIISPAVAGVQONA
VGGAIPVNVKSLPLRAGEETVPRIGTIVRNQALPPSARALGIPRQGVLOS
VEGTPAAGLARGPRSORPPGDIRLGGDITTAVDGGEVSTVODLQVLLGKQSD
EVLILTRDQCTLQRLTLLS"
3113..3118
3127..3678
/note="orfC, proline-rich protein"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4053.1"
/db_xref="GI:7263005"
/translation="MTFNCAAMASRSTRSGRCATRSRSLGVGSESAVPAATV
ARLPVPPVRSAPFLGHDSDMHRPAPSGAARLPVPPVRSVGVGCSAPVLSRLP
RSTPHIRKISFEWRCFSASPLLRGQIVPGRAGGPPRLRPHWRPARACFP
AARPLVSLSCICAVTLVLLSCI"
3691..4038
/note="orfD"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4054.1"
/db_xref="GI:7263006"
/translation="MTTAPLPSASAVGSRGSAVAVSGPGAAWYMCNRRTAAR
ATLRIPASAPSTTAATVAVAMVAGRLRQVRRACGVLMKCSPPRAQIAAAG
WMAVIVATGRRQ"
4090..4551
/note="orfE"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4055.1"
/db_xref="GI:7263007"
/translation="MAAPTRKHSPPPWLLPLAGLMSVAPVALTTPPLPPPCSAI
TTPPPACCPVLTVPRTSPVGRSSTPPSPCFPPRGSAPARBMPCGWSRAA
CPFSLAPLRPPRPARPGRPPQAPFAGSGRRRPPACSSFPAPARLSFI"
4666..4670
4678..5364
/note="orfF"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4056.1"
/db_xref="GI:7263008"
/translation="MSELTFRUSPHARALVALGADHTAPAPARAADIYMOQISGP
APVTDAPRSAPARTLRPLRGVYGRWDSAPLPTARACKLNLTLTAGADPITMAR
ALTRGDTAAEQITCEGMQLDIQNRRAQLAAQOQOQAPAPAPLTLRTQARLHK
QIAGIWTDERVTOAKIARRPVESLITALLNODBAALLVVAABEARDRAAVLAQOD
PAAGCALIR"
5407..5973
/gene="ardu"
5407..5412
/gene="ardu"
5422..5973
/gene="ardu"
/codon_start=1

gene

CDS

1187..1192
/gene="htru"
1198..2436
/gene="htru"
/transl_table=1
/codon_start=1
/product="htru"
/protein_id="AAFA4048.1"
/db_xref="GI:7263000"
/translation="MKTAMPILTLALLAGCTDPESTRQTTSSSRKATTTMPARGA
ARDSSSPGAGTASPDQRLIEDNTIDVMDQGVVYTRDQSGGALFESBP
LAPPAEDQGPFGSGGFLIDREGHILTNVHIDADIRILQNRQDPAVVG
TAPVYDALLRARGARRGAGGARGSPAGRKDCGAPGLEFTYQGLITFAVK
RVIPMGVESLPONSQVTDAAINPCNSGCPVANSRGVGNQIISPAVAGVQONA
VGGAIPVNVKSLPLRAGEETVPRIGTIVRNQALPPSARALGIPRQGVLOS
VEGTPAAGLARGPRSORPPGDIRLGGDITTAVDGGEVSTVODLQVLLGKQSD
EVLILTRDQCTLQRLTLLS"
3113..3118
3127..3678
/note="orfC, proline-rich protein"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4053.1"
/db_xref="GI:7263005"
/translation="MTFNCAAMASRSTRSGRCATRSRSLGVGSESAVPAATV
ARLPVPPVRSAPFLGHDSDMHRPAPSGAARLPVPPVRSVGVGCSAPVLSRLP
RSTPHIRKISFEWRCFSASPLLRGQIVPGRAGGPPRLRPHWRPARACFP
AARPLVSLSCICAVTLVLLSCI"
3691..4038
/note="orfD"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4054.1"
/db_xref="GI:7263006"
/translation="MTTAPLPSASAVGSRGSAVAVSGPGAAWYMCNRRTAAR
ATLRIPASAPSTTAATVAVAMVAGRLRQVRRACGVLMKCSPPRAQIAAAG
WMAVIVATGRRQ"
4090..4551
/note="orfE"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4055.1"
/db_xref="GI:7263007"
/translation="MAAPTRKHSPPPWLLPLAGLMSVAPVALTTPPLPPPCSAI
TTPPPACCPVLTVPRTSPVGRSSTPPSPCFPPRGSAPARBMPCGWSRAA
CPFSLAPLRPPRPARPGRPPQAPFAGSGRRRPPACSSFPAPARLSFI"
4666..4670
4678..5364
/note="orfF"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="hypothetical protein"
/protein_id="AAFA4056.1"
/db_xref="GI:7263008"
/translation="MSELTFRUSPHARALVALGADHTAPAPARAADIYMOQISGP
APVTDAPRSAPARTLRPLRGVYGRWDSAPLPTARACKLNLTLTAGADPITMAR
ALTRGDTAAEQITCEGMQLDIQNRRAQLAAQOQOQAPAPAPLTLRTQARLHK
QIAGIWTDERVTOAKIARRPVESLITALLNODBAALLVVAABEARDRAAVLAQOD
PAAGCALIR"
5407..5973
/gene="ardu"
5407..5412
/gene="ardu"
5422..5973
/gene="ardu"
/codon_start=1

gene

CDS

misc_feature

RBS

CDS

gene

gene
CDS

/translation="MSKKYNPPTKTHKVVSSEIRKSTUCKRCHEMDGRISENLMLRLK RKHEDAKTVLDEKRETRALAOQANAKDVAVNTTKVALYLPTTCKOSITPEDD DSQSHGEATWLTGEEETGVDNKNNAIVSOFPGVTGSEIPDEYOKRNSNMMLK AVDVASVOTNESSVCAPASTARAKSPARRASSAAADLOSDSDIEDTLRTLCAGLV KYSARTNVLGFHINRLTYKADMCHFEETDSRMGDDDYASP"
2911..4557
/gene="EG:BACH48C10.2"
join(2911..3312,3484..4557)
/gene="EG:BACH48C10.2"
/note="prediction:(method:"genefinder", version:"084", score:"52.65");
/prediction=(method:"genscan", version:"1.0", score:"112.75");/match=(desc:"GH19593.Sprine GH Drosophila melanogaster head port2 Drosophila melanogaster cDNA clone GH19593 Sprine, mRNA sequence", species:"Drosophila melanogaster (fruit fly)", ranges:(query:2855..3324,target:EMBL:A1388635:1..470, score:"2296.00"),(query:3478..3538, target:EMBL:A1388635:453..513,score:"287.00")), method:"blastx", version:"1.4.9")"
/codon_start=1
/protein_id="CAB63508.1"
/db_xref="GI:6594144"
/translation="MYTLRFLLPRRIIDYCRRLSLDTLVTLRPFYAIRPDGLAHNP AEKRVNESLPLSHLESLTPGSASCILPBGHLEFI PSQSDSGSRVSRILPVQV DKSILDPRAVICQLLRSCKLPEKEODRGMEELERLKRKEVRERVAPOL PVQNKLSARELFKDKAREHSLEIMPAOLAQSQAVALTGKIEGIALVOULGI RGTRYDCGNMERFYNHCLSIKGSEDPKVAGALRNRYAVPYGISLEDVMIL FSGSDSGMALFLSMRLHEEDRLVPLYEKALSALLIGIRIERRRYPGCCBARA VGLRRIRIVASEHLRLDESDVELPASLDQHILAIYPDIDTKVSYTGFLAPASC GP DLVEFCRNLDVASGFRFSKYRDIOVERQLMQCEDELGLORLSKDSVSTPDKMYAI RHLOGSKSRGLSHHTNVMSVPDGIICIPDMFSREK"
16590..19135
/gene="EG:BACH48C10.3"
join(16590..17843,17908..18048,18117..18490,18556..19135)
/gene="EG:BACH48C10.3"
/note="the EST emb1|A1063630|A1063630 comes from the 5' UTR"
/codon_start=1
/protein_id="CAB72239.1"
/db_xref="GI:6911910"
/translation="MSSSESTGGDSIDLTYPLAEELHNOLVKHWRENIDALNAKFAN LDEPPAYLIGESKAELMTWTYGLAPPSTSKLPFKYTVYVGCSLRMLEPTEOEL TSKLHELAEADELEMLRLNSDOODEBSLYVERKRPYONOTOILOOROLARYHHG NDLTLDISEPOGCCTLRQPIILRLRALPNQORTSVVISGVRCLDALMAKLALRO IFLPMDCFVSTTHSGRIHPWHDIQTLAAEELFRELDRFLPTHKHQITKTFTSL VFPECGRILTFPFYCOCNRFPHQCARVIMLCOPFPDMSYQDLLAENPDNGVF PGCVATREMMSRSRSCRSSSSSSSSSGNHRQHPPI SODDRAADSN VCIINRTSVTSYEORSIMQARPPLPCPDHNSNSTOASPTSLKINRRARSADSN KNLRLRAKSSDENMLIAEELLIGRTSGSGTCYRAHMHGPVAVKTLANKTSPSA QILAQRVAMALKTRHCNTLRFPGCYKPSLAIVYOWGESLKYHVAVSFRTKLN TLIDIGKQVAMLKTRHCNTLRFPGCYKPSLAIVYOWGESLKYHVAVSFRTKLN ANOPTGSLIMAAPEVIMOEINLPYSQSDYAGVIGAYTELACPLGYHHSNDOLF HMGVGLIRPMSOVARBPALRIALDECITKYPKRPFLRLANNLEMLRLPLPXIF RSASEPMLTOSOLONDEFILPSPKTPVNNEPFGSGAGNI"
complement(19754..20239)
/gene="EG:BACH48C10.6"
complement(19754..20239)
/gene="EG:BACH48C10.6"
/note="EG:BACH48C10.6"
/complement(19754..20239)
/gene="EG:BACH48C10.6"
/note="EG:BACH48C10.6";
/prediction=(method:"genscan", version:"1.0", score:"82.07");/match=(desc:"SIMILARITY TO RIBOSOMAL PROTEIN L14", species:"Caenorhabditis elegans", ranges:(query:20027..20128,target:SPTREMBL:020430:36..33, score:"94.00"),(query:19952..20011, target:SPTREMBL:020430:67..48,score:"73.00"), (target:SPTREMBL:020430:109..80, query:19835..19924,target:SPTREMBL:020430:109..80, score:"88.00"), method:"blastx", version:"1.4.9")"
/codon_start=1
/protein_id="CAB63504.1"
/db_xref="GI:6594140"
/translation="WAIRILKTVROVADPLAVTLGGOGTOQLHTTTACEIRIKRLAT

gene
CDS

```
RVVDSNDI.GKAAAEGRPRCIRHYNNRCRGEGTGGDVIVA.IKCOMMKGLTVGKONOK  
PKQPKFDSNNNLVLIDDNGSPDGTIRIHPPIPTILRTILKEKTLAKGADYTKVALASRY  
V"  
complement t (21238 . 25103)  
/gene="BG:BACHA8C10.5"  
complement {join (21238 . 24581, 24649 . 24951, 25046 . 25103)}  
/gene="BG:BACHA8C10.5"  
/comp="BG:BACHA8C10.5"
```

Query Match:	64.0%;	Score 16;	DB 3;	Length 49587;
Best Local Similarity:	100.0%;	Pred. No. 55;		
Matches 16;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	5	gtgcatacgcacatt	20	
Db	48632	GTGCATCGCCAGCATTT	48647	

RVVNDNSD,IGKAAAEGRPRRCIHVYNNKRGVIGCDKVLVAIKGOMKGGILVGIKONK
PKQKRFPSNNLVLIDDGNSPLGTRIHVPILPTILTTLTKETTLAKGADYTVLAISAY
V
complement(21238..251103)
/gene="EG-BACH48C10.5"
complement(join(21238..24581,24649..24951,25046..25103))
/gene="EG-BACH48C10.5"
/note="(prediction(method:"genscan", version:"1.0",
score:"314.45")); (prediction(method:"genefinder",
version:"084")); (match(desc:"GH02860.5prime GH
Drosophila melanogaster head port2 Drosophila melanogaster
Drosophila GH02860 5prime, mRNA sequence",
cDNA clone GH02860 5prime, mRNA sequence",
species:"Drosophila melanogaster (fruit fly)",
ranges:(query:25045..25290),
target:EMBL:A106464.3:369..124, score:"1221.00"),
target:EMBL:A106464.3:369..124, score:"1221.00"),
(query:24786..24952, target:EMBL:A106464.3:355..369,
score:"817.00")); (method:"blast", version:"1.4.9");
match(desc:"LD12603.5prime LD Drosophila melanogaster
embryo Bluescript Drosophila melanogaster cDNA clone
LD12603 5prime, mRNA sequence", species:"Drosophila
melanogaster (fruit fly)", ranges:(query:25040..25317,
target:EMBL:AA43845.3:05..28, score:"1363.00"),
(query:24647..24952, target:EMBL:AA43845.3:04..299,
score:"1422.00")); (query:24549..24579,
target:EMBL:AA43845.3:635..605, score:"155.00"),
method:"blast", version:"1.4.9");
match(desc:"bs1f02.y1 Drosophila melanogaster adult
testis library Drosophila melanogaster cDNA clone bs12f02
5', mRNA sequence", species:"Drosophila melanogaster
(fruit fly)", ranges:(query:23494..24011,
target:EMBL:A194541.3:522..5, score:"2554.00"),
method:"blast", version:"1.4.9");
match(desc:"GM04080.5prime GM Drosophila melanogaster
ovary Bluescript Drosophila melanogaster cDNA clone
GM04080 5prime, mRNA sequence", species:"Drosophila
melanogaster (fruit fly)", ranges:(query:22290..22894,
target:EMBL:AA80224.9:605..1, score:"2971.00"),
method:"blast", version:"1.4.9");
match(desc:"GH25306.5prime GH Drosophila melanogaster
head port2 Drosophila melanogaster cDNA clone GH25306
5prime, mRNA sequence", species:"Drosophila melanogaster
(fruit fly)", ranges:(query:21634..22176,
target:EMBL:A1405337.5:43..1, score:"2706.00"),
method:"blast", version:"1.4.9")
/codon_start=1
/protein_id="CAB72240.1"
/db_xref="GI:5911911"
translation="MSMSKSPFATLNFPGKKRKSCTGNSMGLKSSSTOHLHLS
TDTINLESOLENYNTIPYPAATNAPAEQFRTIVGLPKOLEFARLGAVALSLKLS
LVGNKQKLDADKATFRSVDVSOYSCETSTGTVGLSEILCHKSESSDNDNPNPPL
FORVGNKQKLDADKATFRSVDVSOYSCETSTGTVGLSEILCHKSESSDNDNPNPPL
AKVLAIPPRKRTAPRPQSDVCIQEQSPDIPDIPAGSEPPPRRLRLTKTPAP
SCPLEDNGKSSKSPSDSDGHYATLDLKPILAGSEPPPRRLRLTKTPAP
PDPOGSGNISLSPLEVPPTNPPIPOAPPTITPILAPPAVLAOVNGNSGN
GVSSKVLINRPTRPEPSLGSATGDDDDNRDTSKQADSGIPAPSPSEPPKESQ
QESSEDEDEAIKYNFKLQTSKPPSAQMLKRLTLADIDEEDHNDVQASITNN
GNLAPPSSTSTPMNVDLISPPRPSDQOMKRNQSSPSGLSDIVDELATLIN
QOURLTLKQPDPMNVEISAKPRLANESTITAKSTRTRGRADSRQAKSGGCTV
ASGAGLSHRSSSHVLLKLEONGJGLQDRSSSELSTGESPSLQSEVITLINSRK
NSLAEGSGASVTEQKILKOLESLKEDOSYGVDCVDSISREASQOSPVPKPK
SNPCVETKQSPFRVAGDPPSPVSLKSPQVBOIRPVERSLPVAVSEQAOA
SHUKEITVQKSEPLSESKSFSVLKSPAPKAVGDDPPKVAEVAEESGNSSS
STIPPNPPIAPVPSVNLNRPNNMALEDSQDRDEPASKPLPTTYSQSPSINFA

```

RESULT 29
AC018267/c
LOCUS
DEFINITION AC018267 60573 bp DNA 09-DEC-1999
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in order of
pieces.
ACCESSION AC018267
VERSION AC018267.1 GI:6552924
KEYWORDS HTG: HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 60573)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10214125 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Source
Location/Qualifiers
1..60573
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 16964 a 14192 c 13649 g 15768 t
ORIGIN
Query Match 64.0%; Score 16; DB 2; Length 60573;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 5 gtgcgtgcgcgcgcatt 20
|||||
Db 27002 gtgcattccgcagcatt 26987

RESULT 30
ATAC009895
LOCUS
DEFINITION ATAC009895 82289 bp DNA 24-JAN-2001
Arabidopsis thaliana chromosome III BAC T21p5 genomic sequence,
complete sequence.
ACCESSION AC009895
VERSION AC009895.4 GI:12408719
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 82289)
AUTHORS Lin, X., Kaul, S., Town, C.D., Benito, M., Creasy, T.H., Haas, B.,
Boman, C.L., White, O., Nieman, W.C., Utterback, T.R., Barnstead, M.E.,
Rong, C.M., Koo, H., Fujii, C.Y., Utterback, T.R., and Fraser, C.M.
REFERENCE Arabidopsis thaliana chromosome III BAC T21p5 genomic sequence
Unpublished
2 (bases 1 to 82289)
AUTHORS Lin, X. and Kaul, S.
TITLE Direct Submission
JOURNAL Submitted (04-SEP-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
3 (bases 1 to 82289)
AUTHORS Lin, X.
TITLE Direct Submission
JOURNAL Submitted (24-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
COMMENT On Jan 24, 2001 this sequence version replaced gi:12280866.
Address all correspondence to:
Xiaoying Lin

```

The Institute for Genomic Research
9712 Medical Center Dr.
Rockville, MD 20850, USA
e-mail: xlin@tigr.org
BAC clone T21p5 is from Arabidopsis chromosome III and is near the
molecular marker ml172.
The orientation of the sequence is from 5'6 to 3'7 end of the BAC
clone.

Genes were identified by a combination of three methods: Gene
prediction programs including GENE (available by anonymous ftp
from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
Washington), GenScan (Chris Burge,
http://genome.stanford.edu/~chris/GENSCAN.html), and NetPlantGene
(http://www.cbs.dtu.dk/netplantgene/cbsnetplantgene.html), searches of the
EST database at TIGR (http://www.tigr.org/tdb/at.html).
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity to other proteins are named
Genes without protein or EST similarity, that are predicted by more
than two gene prediction programs over most of their length are
annotated as "hypothetical" proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/SW/RepeatMasker.html). Regions of
genomic sequence that are not annotated as genes but have predicted
exons by GENE are annotated as misc features.

FEATURES

Source

Location/Qualifiers

1..82289

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db_xref="taxon:3702"

/chromosome="III"

/map="ml172"

/clone="T21p5"

1..24217

/note="overlap with BAC clone T12J13
(AC009327:65718..89934)."

complement(join(8..215,451..587,675..857,959..1050,
1459..1593,1682..1748,1837..2053,2146..2340,2492..2879,
2979..3100))

/gene="T21p5.1"

complement(8..>3100)

/gene="T21p5.1"

/note="similar to hypothetical protein GB:CA38918
[Arabidopsis thaliana]"

complement(join(8..215,451..587,675..857,959..1050,
1459..1593,1682..1748,1837..2053,2146..2340,2492..2879,
2979..3100))

/gene="T21p5.1"

/note="unknown protein"

/codon_start=1

/protein_id="AA01600.1"

/db_xref="GI:6017117"

/translation="MSQISGNNNIPLESYVSLVKNADKFSKIRDLPEYRSRYEN
YFVRFVYVYQLMKFDQENRQKVEAGLAKWEIGETIASIAQIYGHVMTSDAGYLS
ESTVEYVAILIFREYFKDGLRQDINANKQIRFLARLAWCLVIGREYVMDQDKR
LIDCKRTFOETDFEKWVAQEIIVELKSDTFNMIRPLRSYLVDPNIDGTPPAS
RSRLRDAIISYYCNEVKSSELTDSFRMLQCLEMEPGSSLYSGACMGONAPGV
ARINSQMSNDPPTLPNDRKAVLYRPSITHTLATICEBPSHGILLIYLSASGKIG
QISSPLSARSATSVENILRDPSHTIKOETPSSQITPSSGYIYPSDLVAFKRP
LFIIDSSSTVKNICGAKGEPALILSPSTPLISADSPROPSGLFTIFLSP
VOAFCLISLISASDMETDIFTKAEKLISSMWMASTLATSPLHVMQSLIKDPLR
RLIRLIFECRAVALIYTPVFNKONQPECCPSLPESLPLTAPVQSAVPMNVGAT
SKFTIPQDITMLESF"

join(3627..3661,3762..3843,3918..3995,4081..4140,
4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,
5361..5494,5576..5682,5785..5827,5903..6144)

/gene="T21p5.2"

<3627..>6144

/gene="T21p5.2"

misc-feature

mRNA

gene

CDS

mRNA
gene

```

CDS
/note="predicted by genscan"
/join(3627..3661,3762..3843,3918..3995,4081..4140,
4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,
5361..5494,5576..5682,5785..5827,5903..6144)
/gene="t21p5.2"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAFO1580.1"
/db_xref="GI:6017097"
/translation="MDNRSSESIKKEHELEKONIASRKLEPTNTKLIDPPEMALYAK
VRSQEEIHSIQLERIAACLDKQMLNKKYGERKCADLRVAIDKONSVTSALNEL
ARRKGELEENIKLAHDAKTEDEERYIFMTSLGLAEYGVWPAVANAITSIGITHLH
DOLQWTKACNDRIRELSTIVENQPGFISKDNHDPNRSKTOASTGSTRDNQOTN
EOLIPMENTRNPIRIMODTESLRFNOIGGSGGIFPOPKRENGYVLSGPEVR
MIOREKKAENSWMFADJNNEFASIVEEGIDGFOIIGDAIPKEKVLGCGPEVR
GTTICMROWARHLEDGCTROYIEGATTHEYIVTADVDKLIIVAEICIPMDQGRQVYRD
FSGIYFSESVSKVDLIMQGLVRLFANDQKIKGYNFNGT"
/misc_feature
6225..6289
/note="exon predicted by xgrall, quality marginal"
6384..6503
/note="exon predicted by xgrall, quality excellent"
6794..6839
/note="exon predicted by xgrall, quality excellent"
6928..7033
/note="exon predicted by xgrall, quality marginal"
7293..7411
/rpt_family="(TAAA)n"
/complement(7488..7547)
/note="exon predicted by xgrall, quality
excellent_shadowexon"
complement(7634..7784)
/rpt_family="(TAA)n"
complement(7819..7864)
/note="exon predicted by xgrall, quality good_shadowexon"
8054..8113
/note="exon predicted by xgrall, quality marginal"
8263..8380
/rpt_family="(TAAA)n"
complement(8592..8627)
/note="exon predicted by xgrall, quality
marginal_shadowexon"
<8652..>9527
/gene="t21p5.3"
<8652..>9527
/gene="t21p5.3"
/note="contains zinc finger motif, C3HC4 type (RING
finger)"
8652..9527
/gene="t21p5.3"
/note="unknown protein"
/codon_start=1
/protein_id="AAFO1602.1"
/db_xref="GI:6017119"
/translation="MGSTGNPNMGTTTYSYRDCSGVCYVCPQWCYVIFPPPSFY
LIDPESSSSSPLILIGILASAFILVSYTILSKYCHRRHNSSTSAAINRI
SSDYWGCTNNNNNGATPNNGTGGGGGIDELSKITVYKRAKMGFEVSSDCS
VCSEFQENESLRLLKRCNHAHFVPCIDPWLKSHSCPCICRAPIYSSAVEIVDLJNO
QIVTENNISTGDDSVVYVNLDENSRSRNEYVNBGSTPTXODAGKRRRRRAELGVK
LRRCGIDRRYTKGD"
/misc_feature
complement(10838..10896)
/note="exon predicted by xgrall, quality
marginal_shadowexon"
1174..11254
/note="exon predicted by xgrall, quality good"
11690..11700
/note="exon predicted by xgrall, quality marginal"
/join(<11814..12256,12381..12578,12676..13085,13162..13332,
13418..13761)
/gene="t21p5.4"
<11814..>13761
/gene="t21p5.4"
/note="similar to unknown protein GB:AC32238 [Arabidopsis
thaliana]"

```

```

CDS
/join(11814..12256,12381..12578,12676..13085,13162..13332,
13418..13761)
/gene="t21p5.4"
/note="unknown protein"
/codon_start=1
/protein_id="AAFO1581.1"
/db_xref="GI:6017098"
/translation="MAETKKGSESEYPIKTIIVLVQENSFHTTLGFKELNREIDGYM
KSNQKRPFGSSDLNSHNVFGDSQYDPNPGHSTIPDYEQVFGKPMDSGHPDNPNG
PATMSGFAONAEKRRKMGSSAVYNGKFPALVYKEIVONFAICGRWASVGAOPN
RLFHSATISGTTNNEKLLIFGPPOKTIYESIDRAGTFPGIYCCFPTLRYRLRK
LKYLIRPHDYGLQPKDKCKEGLNPNVVEQKRTDILLPANDDPRSHDSGCKLYK
EVEALRSPQMMEDILFIYDDEGGRVYHVPPIIDGVNPDGILGPPYNERRLG
VRVPTFTSPWIEGTVLHSGNPGYVLSQGEHSSIPATYKTIFFKLKDELTKRDNAGT
FEVYITRNSPDDCCPETLSNPKRGVAKENALSDFOELVITYAAGIKGDKNEEL
LRIACKTCVSADSASYTKRFDKFVSESKARRRGDENDIVFCVDDDDHNVKPP
S0SEPSHATPMSN"
/misc_feature
complement(14488..14563)
/note="exon predicted by xgrall, quality
excellent_shadowexon"
14893..14935
/note="exon predicted by xgrall, quality marginal"
/join(<15141..15586,15675..15872,16258..16667,16780..16950,
17036..>17427)
Query Match 64.0%; Score 16; DB 8; Length 82289;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 aggtgcacgcagca 18
Db 30302 AGGTGCATCGCCAGCA 30317
RESULT 31
HS272E8/c
LOCUS
DEFINITION
HS272E8 117968 bp DNA PRI 23-NOV-1999
Human DNA sequence from clone 272E8 on chromosome Xp22.13-22.31.
Contains a pseudogene similar to MDM2-Like P53-binding protein
gene. Contains STSS, GSSS and a CA repeat polymorphism, complete
sequence.
293929
VERSION 293929.1 GI:3425887
KEYWORDS HNG; CA repeat polymorphism; MDM2-Like P53-binding protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 117968)
REFERENCE
AUTHORS Grafham,D.
TITLE Direct Submission
JOURNAL Submitted (17-AUG-1998) E-mail enquiries: humquery@sanger.ac.uk
On Aug 18, 1998 this sequence version replaced gi:3334544.
Clone requests: clonerequest@sanger.ac.uk
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
IMPORTANT: This sequence is not the entire insert of clone 272E8.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone 410B11 (286063) is at 117869 in this
sequence. The true right end of clone 390N22 (A0008711) is at 61633
in this sequence.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.

```

This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChX>
2728 is from the library RPCI1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pcypac2>.

FEATURES

source

```
1..117968
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="p22.13-22.31"
/clone="RP1-272E8"
/clone_id="RPCI-1"
1370..1566
repeat_region
/note="MER20 repeat: matches 2..218 of consensus"
complement(3645..3946)
/note="MER33 repeat: matches 324..2 of consensus"
complement(3864..3994)
/note="MER3 repeat: matches 131..1 of consensus"
4194..4555
/note="MLT1A1 repeat: matches 5..365 of consensus"
5433..5531
/note="MIR repeat: matches 87..188 of consensus"
complement(6029..6111)
/note="MIR repeat: matches 188..103 of consensus"
6205..6322
/note="MIR repeat: matches 106..256 of consensus"
7754
misc_feature
8261..9521
/note="predicted CpG island"
complement(9741..9813)
/note="MIR2 repeat: matches 146..74 of consensus"
complement(11533..11605)
/note="MIR2 repeat: matches 145..74 of consensus"
complement(18288..19413)
/note="LIMD2 repeat: matches 1088..5 of consensus"
complement(19266..19525)
/note="L1 repeat: matches 5390..5135 of consensus"
complement(19531..22106)
/note="L1 repeat: matches 4931..2303 of consensus"
complement(22115..22419)
/note="AluX repeat: matches 301..1 of consensus"
complement(22424..22641)
/note="L1 repeat: matches 2314..2093 of consensus"
complement(23516..23641)
/note="MER25 repeat: matches 2036..1908 of consensus"
25175..25396
/note="match: GSS 082600"
27948..28235
repeat_region
/note="AluX repeat: matches 1..301 of consensus"
complement(28665..29020)
/note="THE1B repeat: matches 364..1 of consensus"
complement(29889..30189)
/note="AluJ repeat: matches 302..1 of consensus"
complement(3138..33487)
/note="MER5A repeat: matches 155..4 of consensus"
33121..33328
/note="MIR repeat: matches 63..262 of consensus"
33424..33723
repeat_region
/note="AluX repeat: matches 2..301 of consensus"
35811..36234
repeat_region
/note="L1 repeat: matches 1575..2001 of consensus"
36723..39034
/note="L1 repeat: matches 2121..4429 of consensus"
37893..38188
/note="AluX repeat: matches 1..302 of consensus"
39040..39339
/note="AluX repeat: matches 1..301 of consensus"
39353..40296
/note="L1 repeat: matches 4449..5390 of consensus"
40148..41111
repeat_region
```

```
repeat_region
/note="LIMD1 repeat: matches 1..968 of consensus"
complement(41972..42157)
/note="MLT1F repeat: matches 541..354 of consensus"
complement(43606..43671)
/note="MIR2 repeat: matches 134..71 of consensus"
45479..45585
/note="MER45 repeat: matches 26..133 of consensus"
48242..48312
/note="MER45 repeat: matches 98..176 of consensus"
49604..49826
/note="AluJ repeat: matches 82..301 of consensus"
49936..50472
/note="L1PA13 repeat: matches 41..573 of consensus"
complement(50932..51134)
/note="MER20 repeat: matches 218..18 of consensus"
51378..51596
/note="MER20 repeat: matches 1..215 of consensus"
complement(52293..52552)
/note="MLT1C repeat: matches 466..186 of consensus"
complement(52555..52908)
/note="THE1B repeat: matches 362..1 of consensus"
complement(52914..53071)
/note="MLT1C repeat: matches 182..8 of consensus"
53212..53621
/note="MSTC repeat: matches 1..403 of consensus"
55202..55295
/note="MIR2 repeat: matches 1..94 of consensus"
55296..55504
/note="MER3 repeat: matches 1..209 of consensus"
complement(55783..56684)
/note="L1PA10 repeat: matches 911..1 of consensus"
complement(56534..57144)
/note="L1 repeat: matches 5390..4785 of consensus"
complement(57148..59003)
/note="L1 repeat: matches 4720..2850 of consensus"
59735..59757
misc_feature
61714..61945
/note="match: STS L24834"
complement(63228..63358)
/note="MIR repeat: matches 247..100 of consensus"
complement(64335..64455)
/note="L1ME3 repeat: matches 909..790 of consensus"
complement(65455..65744)
/note="AluJ repeat: matches 289..2 of consensus"
66343..66557
/note="MIR repeat: matches 41..261 of consensus"
66912..67656
/note="L1ME3 repeat: matches 1660..2409 of consensus"
complement(68490..68794)
/note="AluY repeat: matches 301..1 of consensus"
complement(68873..69190)
/note="AluS9 repeat: matches 298..1 of consensus"
69217..69532
/note="MER33 repeat: matches 1..324 of consensus"
complement(69638..69703)
/note="MIR repeat: matches 129..64 of consensus"
71210..71340
/note="MIR repeat: matches 83..206 of consensus"
72426..72792
/note="match: GSS A0053969"
complement(72923..72973)
/note="MIR repeat: matches 139..89 of consensus"
complement(75007..75196)
/note="MER5A repeat: matches 189..1 of consensus"
complement(75386..75425)
/note="L1ME3 repeat: matches 406..367 of consensus"
75508..75800
/note="AluJ repeat: matches 1..288 of consensus"
complement(76744..76840)
/note="MLT1F repeat: matches 182..91 of consensus"
complement(78342..78521)
/note="MIR repeat: matches 216..35 of consensus"
complement(78988..79020)
repeat_region
```

repeat_region /note="MIR2 repeat: matches 144. .112 of consensus"
 /note="MIR2 repeat: matches 134. .85 of consensus"
 repeat_region /note="MIR repeat: matches 242. .3 of consensus"
 repeat_region /note="MIR repeat: matches 178. .5 of consensus"
 repeat_region /note="MIR repeat: matches 109. .27 of consensus"
 repeat_region /note="MIR repeat: matches 11. .91 of consensus"
 repeat_region /note="MIR repeat: matches 141. .87 of consensus"
 repeat_region /note="MIR repeat: matches 252. .177 of consensus"
 repeat_region /note="MIR repeat: matches 147. .29 of consensus"

Query Match 64.0%; Score 16; DB 9; Length 117968;

Best Local Similarity 100.0%; Pred. No. 54;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 atccgacattcg 24

Db 97872 ATCCGACATTTCGA 97857

RESULT 32
 AP002817/c AP002817 140825 bp DNA PLN 12-AUG-2000
 LOCUS Oryza sativa genomic DNA, chromosome 1, PAC clone: P0699D11.
 DEFINITION AP002817 BA000010
 VERSION AP002817.1 GI:9558510

KEYWORDS Oryza sativa (cultivar: Nipponbare) DNA, clone: P0699D11.

SOURCE Oryza sativa
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Eukaryota; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 140825)

AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.

TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC

clone: P0699D11

2 (bases 1 to 140825)

3 (bases 1 to 140825)

4 (bases 1 to 140825)

5 (bases 1 to 140825)

6 (bases 1 to 140825)

7 (bases 1 to 140825)

8 (bases 1 to 140825)

9 (bases 1 to 140825)

10 (bases 1 to 140825)

11 (bases 1 to 140825)

12 (bases 1 to 140825)

13 (bases 1 to 140825)

14 (bases 1 to 140825)

15 (bases 1 to 140825)

16 (bases 1 to 140825)

17 (bases 1 to 140825)

18 (bases 1 to 140825)

19 (bases 1 to 140825)

20 (bases 1 to 140825)

21 (bases 1 to 140825)

22 (bases 1 to 140825)

23 (bases 1 to 140825)

24 (bases 1 to 140825)

25 (bases 1 to 140825)

26 (bases 1 to 140825)

27 (bases 1 to 140825)

28 (bases 1 to 140825)

29 (bases 1 to 140825)

30 (bases 1 to 140825)

31 (bases 1 to 140825)

32 (bases 1 to 140825)

33 (bases 1 to 140825)

/cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /chromosome="1"
 /clone="P0699D11"
 join(527.679,925.1114,2076.2467,3067.3555)
 /note="EST C22619(s11214) corresponds to a region of the
 predicted gene.
 Similar to Arabidopsis thaliana chromosome 2, BAC clone
 T17D12; putative protein kinase (AC006587)"
 /codon_start=1
 /protein_id="BAB03429.1"
 /db_xref="GI:9558511"
 /translation="MVCIRKNGGCHATITFLRELADATNMFSTECCLGEGGVSIVK
 APLNDQVAVKQDLINGLOGNREPLVEYMLSLHNPVLKLFYGVYDGDRLIYE
 YMLGSLDELDHDLRPGQEPDLMTFMKTLAADAALLEYLHDEALPAVYRDIKPSNI
 LLSGYNALSDDELAKLGVGCKTIVTVMYKTHGCAPEYLSGKTLIKSDIYFSG
 VVLELITGRALDSNRPEDQDVAMARLFLDQKPKMKDPSLHGHFPRGLFOA
 LAIAMCLOEKAKNRPISREVAVASLASOTHEQNTAKRTLLPQSPVRYLDQIN
 QDTSLSQHGVMHPLAGTDHVMQVEYKENCSSSRHPRGKRVTPNGADKERALAEANV
 VWEAMRQEKTSKMR"
 complement(8769.9326)
 /note="ESTs AU093260(C52190), AU062991(C52190) correspond
 to a region of the predicted gene.
 hypothetical protein"
 /codon_start=1
 /protein_id="BAB03430.1"
 /db_xref="GI:9558512"
 /translation="MDOFLPEPTLELCRAFIIRGISDRIKVALSKEPALLVDSAFMSI
 AVDOOSCYMDICDRGNGKIFPLDVGSSDPHISDIAIDCLVLLKAPREAVTGLTN
 LPLKSLVDLKLHGTCLRLTLGLCLAFSCROYLAQGMIVISLSALMREVALYSAFK
 GSHDCIADAAVSYLGAELQRLPRG"
 complement(join(15587.15727,15868.15915))
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAB03431.1"
 /db_xref="GI:9558513"
 /translation="MNNFHYEALIGRCHSVYVYGRKKKSTEFYFAVKSVDKQSKVL
 NEVSAPSSPCSFISSW"
 complement(join(18733.18801,19343.19489,19843.19905,
 20016.20087,20611.20691,20769.20819,20976.21072,
 21611.21753))
 /note="ESTs C72327(E1437), AU07852(E1437) correspond to a
 region of the predicted gene.
 Similar to Archaeoglobus fulgidus competence-damage
 protein (AE000949)"
 /codon_start=1
 /protein_id="BAB03432.1"
 /db_xref="GI:9558514"
 /translation="MISDRILFRAGRTKKNISVSSNGNSNEVEBOTISRSASIIYV
 DELIFGTVDKLGAGICLKLALIGMVSIVAVSNEDISVAEVEYRCKSTPDMEYLG
 GLPLHSDISLAGVAKAFGVRLAPDEEVEYLSLIGNYTGDRREKALPEGLTELL
 HHMKLPLIKCNVILATNATVDELETEMGCLDTQSGLVMAKSFVSKHLCTSLD
 VKIAPVAKICIDFSDVYGMV"
 join(30551.30609,30919.31139,31223.31578)
 /note="Similar to Arabidopsis thaliana chromosome 4, BAC
 clone F10M23; hypothetical protein (AU035440)"
 /codon_start=1
 /protein_id="BAB03433.1"
 /db_xref="GI:9558515"
 /translation="MLCEGRGRREGEENEDIVAKKSMGPDPLPHLRGKSIKVP
 OLIVGIGKNVEEPDNRVAGSLNCKTRGSSVVLPLVLYVAEEMNPRHVRKGNH
 GHDLRQAGAAVARRAAHSSVLEAGSMKGCDDHVRNVAVLRTGLD"
 GDAWVSPKRVARRAAHSSVLEAGSMKGCDDHVRNVAVLRTGLD"
 complement(join(31912.32455,32698.33684,34091.34341))
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAB03434.1"
 /db_xref="GI:9558516"
 /translation="MKKKRNTLLIVGEATFFFEERCRRRAGVSEVERCDKSKP
 SRGRHPRATRAAPLRRBOKRARDEDLATSSGWTPESGVPHNPLPSSPLISLH
 SLLIFSGRLDEHGGEAASAGSVASGSGGAAAAAAGGCGGCGGAPATPS
 RHPIASAAAPGARDADAGLVGGVGGGGEAGGGAANSRRIOIGDERHMKRKA
 PSV VCCPPTFVALVLAIVCPAALCHLAVRGHRRVCSAKOKEMGELLTLTYTSPRS

JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

FEATURES
 source

Location/Qualifiers
 1. 140825
 /organism="Oryza sativa"

CDS

CDS

CDS

CDS

CDS

```

LMAAAKARKVEEPPATPMAEHLGEAEVAAWKSGLQALHNVARLLSAGSEDAAPAA
DLKVLPLVLAELPVLPLRLPRHVRSLSCSLPRKLASTLSPORORNDOSKHTT
TEGNKTKERRENSIIAOKETCRSSAPATPPTVAVARATPNDNSRWGAPASAK
FEQGGGGGDDAALPACRLAAYSLASAPKVEPKPAPRPTJTNHPIKORAPQCR
SSPSLSSSSSLSSGKAPRCGVAARELLPMRRKOLRATSIHAEARAVLVIVL
VKLPAPKRECPKSTAPSAKPPHRR.
Join(35382..35472,36089..36172,36426..36530,37078..37337,
38995..39114)
/Note="hypothetical protein"
/codon_start=1
/protein_id="BAB03435.1"
/db_xref="GI:9558517"
/translation="MOLDASDPHGOELIKLHROCKLSLESSSSKSYDLVSPAAQOP
DATVFLPSGSGKAVGKTARVSLVITPDPELEFPGGLSTPPSSKSGASGSKTNR
PTSGSKNRSOSSAOSLQOQSAEPFAHKNOKKDESSLGHNSVSKSAGEQVQGDS
GTTSEVPOTPEKSKKKKKLDRPWKRRTPYDRHKTAQELDRTAVVDSLYKSPKIVY
Q."
Join(40731..40753,43157..43538,45516..45821,47347..47893,
48657..48934,48983..50395)
/Note="Similar to Arabidopsis thaliana chromosome II BAC
clone T27A16; hypothetical protein (Ac005496)"
/codon_start=1
/protein_id="BAB03436.1"
/db_xref="GI:9558518"
/translation="MYSLRPKIRPPPALAARDSPLRWPPPPGRLSPWTRKGR
RRLIPYAMSRGPPESGORGRENEEERSERGEGGDDGEGEIKMGREKREKE
EIRLQAGMRDAGCTGSPADFLSPNNKSLVSLAKDEIKOLACNRRRDPDGLTG
HGAAKDPLDIRSTYEAKLHAGSTIHLASCSWHLHPCSSCKAKGLARRSKASASAL
LSSGSMNPPTOKCPKNSORPATTLLTISLISLIPPELSPHNSYSCIEFISI
SLNDFGRFRPPPTSTIRPTAMRYVRSIVIPVDGYEYERKILAPYTSISMSS
SLNMEQTVKICGSALSLGIGTPIHRAKPIAFPGYCNPFELPCKTEALAH
VYVHSSINRAKPIKIDILYKQAPNLTSPCRNRRHTHSHOCHTHRLPQ
YRVLSEAHKQREARETARPHDRTSIEBNRPAPATVAREHPSVSLRAAARLNG
DLRAHAYLLPGRGHRHRAVASALLSLRPLRDAHALVRRHPSVSLRAAARLNG
RGCGGLAOLSHLLVRAGHADPASAALVQACSGSVASARVDEITASADVVSW
NMIDGYVSGDILARARELFDYMGGRNVYRNTVIGAYAOVKQPEEALIEVRMOYEG
IEPDVALLSVISLSCGLGVYDGEVHRELVKRGICQELPLMNLIDMYKCGSVK
ALPYFGECKSTVWTMTAGLHGLAEVLEPRMKEVNSPNDITFLAVLSVC
SHVGLTDLGMYERTWYSQYIKPRVEHYGCMIDLLQRCLEAGLLIDMPFKANA
ALWQALLAARHODTELGQALHLELEPHNSCNVILISNTYARQEDRDAVREIKI
SMRDRGLRNVGASSIDVGMVHEFTSRDGSHPSLKIREVLCAINSKSVHTAILI
PESLHVEEG."
complement(join(52441..53547,53887..54014,54128..54165,
54272..54487,54709..54870,55838..56016,56150..56256,
56375..56539,56658..56741,56821..57040))
/Note="Similar to Arabidopsis thaliana chromosome I BAC
clone T5A14; kinesin (Ac005223)"
/codon_start=1
/protein_id="BAB03437.1"
/db_xref="GI:9558519"
/translation="MDVQPARTMRNLPTLSSLMGFNKLTPSWIESVSHIDGLSPY
KIDMKYKVEKENDISDNTSEKAVKQIDDELVSILMAQIKOITLQREALNNYLDK
NIRVFCRIPIPHIEESYSRNLFTIDESVFLKVAETKRYQKPKVFOFSTODGF
SEVEYFKALDQYVNCIFAYGQSGKTYMEGKPTNIDPGICQITFNQASNSP
RELETFMSLEIYMGNIJDLAPRSKTNIGKNVPSIKSDPGQIEEDIVAVYNSP
QEVKRLYEMGTRLRSTASTMANSTSRSHCLIRISLTINATERKATSKLMNIDLG
SEVLYTKATGRKIRKGAINTLSALGVADILQTKPHVYRNSKLTQVLRDLCG
NLEHEDGIRRKIKULESMELHLPQPTIYNFDMSHLSEELKTQVSSVNRKN
REBASRLPRFKPTASSOHRIGLNNRPDIINRLPPVPRRPPSSVAESVYVNA
AWQSECSSECSMSLSDMNWTPSIDRGEGCSMDLSEELKTQVSSVNRKN
YTDYDLAESRDIOIKIEKGIVDIDNMLHOOVETSTFRSKMTIDITGVBAEHVS
SIPSTPMAGCTKEDSOVKDEWGLTLOSSTDVEDIKOSKTDNQTAAELCTPPEKEP
SSNNEVKGKHNHPYHGRPRSLDQELLENCTLEPNMDSKSHRSHDKHKTGNVLSY
TGE."
Join(57603..57645,57742..57843,58195..58344,59425..59479,
59537..59589,61304..61785)
/Note="hypothetical protein"
/codon_start=1
/protein_id="BAB03438.1"
/db_xref="GI:9558520"
/translation="MKQDKGNRCLFLQGSFAPYIHQHVITKGECPDTPPSRDR
SEYAKLMTDITPMHMTTFNGMRCLEANENHSRAVIRKSGPTMCHAELESEATAIS

```

```

CDS
/Note="Similar to Arabidopsis thaliana DNA chromosome 4,
BAC clone M3E9; hypothetical protein (AL022223)"
/codon_start=1
/protein_id="BAB03439.1"
/db_xref="GI:9558521"

```

```

Query Match 64.0% Score 16; DB 8; Length 140825;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 8 catcgccagcatctcg 23
|||||
CATCGCCAGCATCTCG 89607
RESULT 33
LOCUS AP001366/c
DEFINITION Oryza sativa genomic DNA, chromosome 1, PAC clone: P0469E09.
ACCESSION AP001366 BA000010
VERSION AP001366.1 GI:7228436
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (cultivar: Nipponbare) DNA, clone: P0469E09.
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 146081)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone: P0469E09
Published only in Database (2000) In press
2 (bases 1 to 146081)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission
Submitted (09-MAR-2000) to the DDBJ/EMBL/GenBank databases. Takuji
Sasaki, National Institute of Agricultural Resources, Rice Genome
Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
Tel: 81-298-38-7441, Fax: 81-298-38-7468)
The orientation of the sequence is from SP6 to T7 of the PAC clone.
Genes were predicted from the integrated results of the
following: GENSCAN 1.0, BLASTN 2.0, BLASTX 2.0 as well as
SplicePredictor (October 1998 version). The genomic sequence was
searched against the non-redundant database NRP (PIR, SWISSPROT,
RCP, Protein Similarities of the coding regions were searched
against NRP with BLASTN 2.0. ESTs represent the identified cDNA
sequences using BLASTN 2.0 with the corresponding DDBJ accession
no. and RCP clone ID.
Detailed information on overlap and assembly quality together with
annotation of this entry at
http://www.dna.affrc.go.jp:82/genomicdata/genomefinished.html.
Location/Qualifiers
1. 146081
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="P0469E09"
Join(5902..5980,6229..6272,8134..8295)
/Note="hypothetical protein"
/codon_start=1
/protein_id="BAA92397.1"
/db_xref="GI:7228437"
/translation="MGAVKRSSFLLVVFALLLTSMAGGRKMLINKHGVQSMETS
DDSMHGGQGEDEMLAMVHERILRQVKTNDYGTPTPMKPIHAKELPN"
9673..9989

```

```

FEATURES
Source

```

```

CDS

```

```

LTR

```



```

CDS
/note="5' LTR"
join(10528..12588,12673..13083,13175..13473,13550..13808,
13901..14083,14241..14306,14743..15045,15216..15479)
/note="EST C28952(C62945) corresponds to a region of the
predicted gene.
Similar to maize transposon MDR mudra protein isolog
(AC003981)"
/codon_start=1
/protein_id="BAA92398.1"
/db_xref="GI:7228438"
/translation="MSSKIDRPAERTFOSIYSLMLRGRIDQEVYTMVSIVVSRAT
EGYFELMMDSTAMRRYVEMAFERSMPLVIFVSVEQDINVMQTEDEGPIINGD
VGVGMONEEDPREBOAMGMDAGEVGLITVDEMEREDINDEADODASSDEGDM
ATDMANEDFSLVILESDHVPMEKXENVDGARYAKHDKMFAVKHMAVSLGREFV
VASTMYEVACMEDECPWRYAHAKGMNDYKWSIYTERKCILOGEYKHNRTTSF
VSTMGYRGQILTAIGDGNQVVPMAFAVESENTEPWFLEHVAIAVRMPNVC
LPRLATIAORNNNTYVDLHFTSVSDRTKSVLORAFSLGACINAVHCRVPLCIDG
TEMTGKRGQILTAIGDGNQVVPMAFAVESENTEPWFLEHVAIAVRMPNVC
LIRHAGMLRAIDYLONGMEKRLPAKMPVRSRCMRHNGANFYKOFKHLIMELF
KRLCAKNOEKFENELMDKLDLTKOTDQSRROVEDEPPIPLGALHDDPTMRRR
SGSSIRNFTOMIENEPIEKMSLLEDDTQSRGIMTTLALAYNNMAGVRLPLVALI
EPIHGTQAVFRDQYKIGPSMADNNIVFGEDVWRLLPHGPAITLPPYSTVGSKAM
LLOFTADLHPDADNNVSRSLAYLLMLPGVMTSTHGAVDPLVHARSIVDQO
PODVOMSGSAVLAATYRALCEACTIDAGAILAGCPMLQJLMAERFAIGRPVDS
APYGVRSQAWPDEGPTMGTYCWRGRRYAHVOYRGYDPFEFDRLOPSDVIMEY
TEBAVAAPRLGLSLCTRODAYMLTLLPMVFDLFEVPHCPORVMROGLOVPFGVY
QPYVLPADHSLTRRQOLAGALWAPVOOYVDDVPLATEEYINELPHTEENYRDYLRN
YLPRTRAVETTPPAPEPVAVIDAPTHDRDTFGADAARDISADITAVOYRLNR
GLHLDVEQRYTFIDMEKRAVMKVFESCSAVIDPAGDPLVCLARPLASECCDLQ
HREFTYRPMCNFTIEMVDENQDGTATYRPSRESDYLRKDEHECRITAAELMHI
VNPGLPTWRREPCRGDCHVITRSVRRRGRCFPCNIVDDDFEFTIHAMMEYER
RVELASIDMRNNPLGLPKRSRPRKCGCDNRQVATSFDEQFRFRYFCFVVDNFVY
CLTYKIRIDHKITTE"
LTR
16135..16469
/note="3' LTR"
join(23418..23748,25558..25589,25715..26480,26838..26965)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92399.1"
/db_xref="GI:7228439"
/translation="MGANCCIAAKERPOPCVTPIVSAFARNVRHSFSPSRDNTHTI
EDIMEPALFENHSSGISIRPEKSGSIAFTDGFSSGSDMFENLKCHKSDRKESSE
KIASRLDRSTTSSNSPEAKLSRSLDTVASDSKMSISVSPTPAISRADPESSE
RGHSLPTDQSRKARSPGOYLOVROVDSKIPSLRSLNEGASPEGRSSMLSYCSN
DISAYSGHSSSDGWSMRTFSFMAVASSOSELGSSKMTSKMSNNPTTH
SPDQECYKCLILKEKSTNNQOELAVANVILCGVYHADCLDSLTAEDKXDPPCPV
STDEGCTVYKFLGLESKTKNKIPKNVIVDNLDDSSKHOKKTIPLBELYSQAASM
RELSKSLGANDMTWLLNDLIEGFS"
CDS
complement(join(27953..28624,29565..29918,30369..30646,
31026..31191))
/note="ESTs C96615(C10106),C26336(C12127),D21959(C10106)
correspond to a region of the predicted gene.
Similar to NAM (AL021889)"
/codon_start=1
/protein_id="BAA92400.1"
/db_xref="GI:7228440"
/translation="MSSLDMVLPQGFPHPCFPHDTELISHYLKKKIHGOKIEYEIPEV
DIYKHPMDLPACQDYPQDNKNHMAARDKRYKGRSRNATVAGYMSGDKRAIK
MGKOTIGTKKTLFHEGRPTGRTGRTWIMHEVYIDRECOACPDMDKDAVLCIRIKRN
DWTPNGNELDNDSDPPEVDAPPVISTEQLPASAEPVAGVGAALPYTAEPDGYTS
AIVANIPSPSDINLDOWNELDFPDEQSLASDLSIDONNESSVCAIARVE
OYSSPENNVDDTEYLPEVDYNIILHPTDFNMIONLPDQPIQVADWVGIOKE
ELMSPOANAEVDSNEADNGILIRYRSKMTPTSVPOKFGTOAKMRAVINKMASS
SFSTNOTIKFENGRLVEHOKNOADVASTKRRDACKPSNELSSNGFLRGIJRNFAAG
CSARMMMLVAFATGAVVAVLHIGORCLSDROQH"
CDS
complement(36747..38998)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92401.1"
/db_xref="GI:7228441"
/translation="MMGHVPLRADTARPMNMMAVPGPEVHVDRHGTPEVPLVLCRPD
DRGPCRASIMAMKMLGILEVHGECCTDDPEHGRPSGTSI"
LTR
39726

CDS
/note="3' LTR"
complement(join(40209..40396,40917..43545))
/note="EST C28952(C62945) corresponds to a region of the
predicted gene.
Similar to maize transposon MDR mudra protein isolog.
(AC003981)"
/codon_start=1
/protein_id="BAA92402.1"
/db_xref="GI:7228442"
/translation="MSSKVTFOIVHGEQNIIRPGDQVDSLDFVTSKIDRPAERTQ
SIYSMLRGFRIDQEVYTMVSIVVSRATGEGFWMELMPSDIDARRRVEMAFESNP
LIVFVSQKODINQNOTDVEGPIINAGVGVPSMONEENOPROARVEMAFESNP
LIVDEREDSDOEEVDODASDEBDDVATPMANDEBSGILSRGHVPPYKERNVY
IBGARVAHDEKCAKMAVSLQREFRVKSTNYVETSRCKECCPWRVYAKKKN
DWKVSIVTEHCEYLOGEYKHNRTISAFMASEMSSVGNIGFEPSITRIENKFK
YITVSVAKRAKOKITEMRGTEFASVDNLPILAIQORNNNTYVDLHFTSVSDRT
KSVLQAFPSLGCINAFHCRVLCIDTFMTGKRGQILTAIGDGNQVLPALA
FVSESTESMYNLELVNIAVVRMPNELLHIDRAGMLRAIDYLONGMEKRLPAK
PDRSRKMEHMKANTYKOFKHLIMELFKRLCAONOEKKEKMLMDKLDLTKOTD
QSRRLPVEGDEPPIPLGALHDDPTMRRSGSSIRNFTOMIENEKKSLLPDDGS
RYGIMTTLALAYNNMAGVRLPLVALIIEFLHGTQAVFRDQYKIGSMADNNIVF
GNVYTKYMEKIKKARRHRYVAGTGVHREYIHCYDRSRGIYRQAOECVAKADG
CTTSCMKPKLHNLPSCHVLAAGDGCIPNNVYSNFRKEALFHTWSEBIVYFGISGS
YTLTSQVPTIPRPSKRLVYKGRROTRRIRNDDESEAGRTLRCSKCDLRGTYKCC
PKRAEVSQADASPSCASDVLTIHVLEQMLLTHHFRCTVSDPARVYHNSHE
KROLELSLTFGAAYLHKVNAHA"
CDS
complement(43999..44385)
/note="5' LTR"
join(44572..44723,44846..44938,46366..46378)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92403.1"
/db_xref="GI:7228443"
/translation="MALPRGVSRPKPMAAASASBPORPGHMLVGTGFGVGRVSO
RLAOGVRSVGTGSPKPKTEMLMDGASIFPATSSRCVW"
CDS
complement(join(48119..48257,48429..48703,49975..50139))
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92404.1"
/db_xref="GI:7228444"
/translation="MALVNDKDEDDVNAEACARRAPIVIGAPMGFVLOARRIGEC
APLPRARKMAITISFCALSLFSEKNTIYILVLCYLRKNRPILVIESKRTCCRTT
ASLARSQDELIVICRARILASSALDRVFLITLSPDLVWLKRCICVINVHISRTH
GSSWSTLPDAVPRKSTREIRLAGIVDTCTGG"
CDS
complement(join(51918..52037,52864..53346))
/note="Similar to Arabidopsis thaliana chromosome 1 YAC
YUPH12R sequence, C2-HC type zinc finger protein
C.e-Myt1. (AC002986)"
/codon_start=1
/protein_id="BAA92405.1"
/db_xref="GI:7228445"
/translation="MASCINGELHVRVYHRRLVKASDESITRPVILAVSNDLIPRTI
QVSMCIYKPSGTGDFHDVVAFAAGIPSLNHFPLAGTIVSNPSCGLPEIHCHQO
GAELVYGEADVALLASLDYGVASGVKLLPEADVLAISOVVSFACGCTEYMGCTNH
VVVGSALSMVLSAMSELAIRSGTILAAAGRPHDSVPRSPSPSYGASLDEAFTPLDG
ARQADARLDEQARATVQAVSAVYIKKRLAAYVGSDBARBMVWVYGRRLTLSS
PELRAMSVYGVNTTAVATAMEIORKLALAYASMARDAIAPVYGEFQELTYM
VEHRAGQORYIDVAVSGVSPASVAFASVDFVTDPGHGAAMALPTSSSARLCT
GFVOIARPPGGGDSWIASALLMRLAALLESDBRLGIRKPVAYEVLGLPAGSSSA
KRAGITSTKI"
CDS
complement(57570..58009)
/note="3' LTR"
complement(join(58018..58891,58980..62098))
/note="Similar to Arabidopsis thaliana chromosome II BAC
F9013 genomic sequence, putative retroelement pol
polyprotein. (AC006248)"
/codon_start=1
/protein_id="BAA92406.1"
/db_xref="GI:7228446"
/translation="NRVNLQOGLTATIDPCYAFRRDMAALSAIIQAVPREMLRGLA
KHDTAKAMDAIKTRVGVDRREKKEGFRQFESNFKRRETPDEARMILTVAVD
IRDMGVMEDEHVNKLRLVYKRYKPAISLEDDLLTKALBELVGLRSTVSYSD
DEGSDGKGLYTEEOMQARVYKOREDSGNSGNKRGAPGTQNHRRKPGGSPKKEA
```

Query Match 64.0%; Score 16; DB 8; Length 146081;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 8193 CATGCCGACGATTTCG 8178
 8 catgccgacgatttcg 23
 |||

RESULT 34
 AC034114/c 154306 bp DNA 09-MAY-2001
 LOCUS Homo sapiens clone Rp11-775E10, WORKING DRAFT SEQUENCE, 8 unordered
 DEFINITION pieces.
 AC034114
 AC034114.4 GI:12958060
 HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFILLTOP.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Graniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 154306)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens, clone Rp11-775E10
 Unpublished
 2 (bases 1 to 154306)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, C.,
 Campopiano, A., Castle, A., Choquel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., DeArrelano, K., Dewar, K., Diaz, J., S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Garand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J., C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Labocque, K., Lamazares, R., Landers, T., Lehotzky, J.,
 Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N.,
 McCarthy, M., McEwan, P., McGuire, A., McKernan, K., McPherson, R.,
 Meldrum, J., Menius, L., Minova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
 Pisanic, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rotman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testafaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (04-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Feb 17, 2001 this sequence version replaced g1:8781909.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/MIT Center for Genome Research
 Center code: WITR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L4353
 Center clone name: 775.F.10
 Summary Statistics
 Sequencing vector: M13; M7815; 41% of reads
 Chemistry: Dye-terminator; Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 151316 bases at least Q40
 Consensus quality: 152394 bases at least Q30

Consensus quality: 153049 bases at least Q20
 Insert size: 151000; agarose-fp
 Insert size: 153606; sum-of-contigs
 Quality coverage: 10.6 in Q20 bases; agarose-fp
 Quality coverage: 10.4 in Q20.
 * NOTE: This is a 'working draft' sequence. It currently
 consists of 8 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 41084: contig of 41084 bp in length
 * 41085 41184: gap of 100 bp
 * 41185 49583: contig of 8405 bp in length
 * 49590 49689: gap of 100 bp
 * 49690 58448: contig of 8759 bp in length
 * 58449 58548: gap of 100 bp
 * 58549 70162: contig of 11614 bp in length
 * 70163 70262: gap of 100 bp
 * 70263 85680: contig of 15418 bp in length
 * 85681 85780: gap of 100 bp
 * 85781 106804: contig of 21024 bp in length
 * 106805 106904: gap of 100 bp
 * 106905 139784: contig of 32880 bp in length
 * 139785 139884: gap of 100 bp
 * 139885 154306: contig of 14422 bp in length.
 Location/Qualifiers

FEATURES
 source
 1. 154306
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Rp11-775E10"
 /clone_lib="RPC1-11 Human Male BAC"
 1. 41084
 /note="assembly-fragment"
 clone_end:Sp6
 vector_side:left"
 41185. 49589
 /note="assembly-fragment"
 49690. 58448
 /note="assembly-fragment"
 58549. 70162
 /note="assembly-fragment"
 70263. 85680
 /note="assembly-fragment"
 85781. 106804
 /note="assembly-fragment"
 106905. 139784
 /note="assembly-fragment"
 139885. 154306
 /note="assembly-fragment"
 clone_end:T7
 vector_side:right"
 BASE COUNT 43533 a 31887 c 31913 g 46271 t 702 others
 ORIGIN

Query Match 64.0%; Score 16; DB 2; Length 154306;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 78184 TGCATGCCGACGATT 78169
 6 tgcatgccgacgatt 21
 |||

RESULT 35
 AC002457/c 177380 bp DNA 21-DEC-1999
 LOCUS Homo sapiens BAC clone CTB-60P12 from Tg21, complete sequence.
 DEFINITION AC002457
 AC002457.1 GI:3947433

repeatL_region	20817. .21123	/rpt_family="Alu"
repeatL_region	21124. .21134	/rpt_family="AT_rich"
repeatL_region	21136. .21171	/rpt_family="(CA)n"
repeatL_region	21478. .21940	/rpt_family="L1"
repeatL_region	22310. .23003	/rpt_family="L1"
repeatL_region	23135. .23653	/rpt_family="MER21_g"
repeatL_region	23658. .24140	/rpt_family="L1"
repeatL_region	24263. .24304	/rpt_family="AT_rich"
repeatL_region	24419. .24447	/rpt_family="AT_rich"
repeatL_region	24618. .24865	/rpt_family="Alu"
repeatL_region	25627. .25650	/rpt_family="(TTAA)n"
repeatL_region	25906. .26152	/rpt_family="L2"
repeatL_region	26153. .26536	/rpt_family="AT_rich"
repeatL_region	26554. .27556	/rpt_family="Retroviral"
repeatL_region	28842. .28903	/rpt_family="(CAT)n"
repeatL_region	29888. .30080	/rpt_family="MER1_type"
repeatL_region	30382. .30659	/rpt_family="Alu"
repeatL_region	32790. .32960	/rpt_family="MTR"
repeatL_region	33062. .33192	/rpt_family="L1"
repeatL_region	33199. .33998	/rpt_family="L1"
repeatL_region	34000. .34042	/rpt_family="(AAAA)n"
repeatL_region	34043. .34316	/rpt_family="Alu"
repeatL_region	34414. .34602	/rpt_family="L1"
repeatL_region	35064. .35263	/rpt_family="L1"
repeatL_region	36253. .36571	/rpt_family="Alu"

Query Match	64.0%	Score 16	DB 9	Length 177380
Best Local Similarity	100.0%	Pred. No. 54		
Matches 16	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	3	agctgcatcgccacga	18	
Db	48179	AGGTGTCATCGCCACGA	48164	

RESULT	36
AC084136/c	
LOCUS	AC084136 180700 bp DNA HTG 14-FEB-2001
DEFINITION	Homo sapiens clone RPJ3-506M4, WORKING DRAFT SEQUENCE, 32 unordered pieces.
ACCESSION	AC084136
VERSION	AC084136.2 GI:12830241
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 180700)

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE	Direct Submission
JOURNAL	Submitted (13-OCV-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On Feb 14, 2001 this sequence version replaced qi:10800269.

Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens, clone RP13-506M4
 2 (bases 1 to 180700)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Barra, N., Bastien, V., Bede, F., Boguslavsky, L.,
 Bouckgealter, B., Brown, A., Burkett, G., Campolongo, A., Castle, A.,
 Choeplak, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
 Deatellano, K., Dewar, K., Diaz, J. S., Dodge, S., Ferreira, P.,
 Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M.,
 Gravel, L., Grand-Pierre, N., Hagos, B., Healdorf, A., Horton, L.,
 Iliev, I., Johnson, R., Jones, C., Kann, L., Kartasas, A., LaRoque, K.,
 Lamanazades, R., Landers, T., Lehocsky, J., Levine, R., Lieu, C., Liu, G.,
 Macdonald, P., Marquis, N., McCarthy, M., McLean, P., McKernan, K.,
 McPheeters, R., Meldrum, J., Meneses, L., Mihova, T., Mlenga, V.,
 Morrow, J., Murphy, T., Naylor, J., Notman, C. H., O'Connor, T.,
 O'Donnell, P., O'Neill, D., Olivar, T. M., Oliver, J., Peterson, K.,
 Pierre, N., Pisanal, C., Pollara, V., Raymond, C., Rieback, M., Riley, R.,
 Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,
 Soungue, C., Spencer, B., Strange-Thomann, N., Stojanovic, N.,
 Strausac, N., Subramanian, A., Talamas, J., Testafae, S., Theodore, J.,
 Tirelli, A., Travers, M., Trifoglio, J., Vassiliev, H., Viel, R., Vo, A.,
 Wilson, B., Wu, X., Wyman, D., Ye, M. J., Young, G., Zaimoun, J.,
 Zimmer, A. and Zody, M.

----- Genome Center
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: W1BR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu

```

----- Project Information
Center project name: L11406
Center clone name: 506_N_4

```

```
----- Summary Statistics -----
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Ptipipe; Build Date 0_960731
Consensus quality: 167122 bases at least Q40
Consensus quality: 174060 bases at least Q30
Consensus quality: 176531 bases at least Q20
Insert size: 18300; agarose-fp
Insert size: 177600; sum-of-configs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 3.8 in Q20 bases; sum-of-configs
```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 configs. The true order of the record
* is not known and their order in this sequence record is
* arbitrary. Gaps between the configs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	952	1051:	contig of 951 bp in length
*	1052		gap of 100 bp
*	1053	2064:	contig of 1013 bp in length
*	2055	2164:	gap of 100 bp
*	2165	3416:	contig of 1252 bp in length
*	3417	3516:	gap of 100 bp
*	3517	4816:	contig of 1300 bp in length
*	4817	4916:	gap of 100 bp
*	4917	6297:	contig of 1381 bp in length
*	6298	6397:	gap of 100 bp
*	6398	9144:	contig of 2747 bp in length
*	9145	9244:	gap of 100 bp
*	9245	11576:	contig of 2334 bp in length
*	11579	11678:	gap of 100 bp

```

* 11679 30905: contig of 19227 bp in length
* 30906 31005: gap of 100 bp
* 31006 32606: contig of 1601 bp in length
* 32607 32706: gap of 100 bp
* 32707 34256: contig of 1550 bp in length
* 34257 34356: gap of 100 bp
* 34357 37618: contig of 3262 bp in length
* 37619 37718: gap of 100 bp
* 37719 39131: contig of 1413 bp in length
* 39132 39231: gap of 100 bp
* 39232 41291: contig of 2060 bp in length
* 41292 41391: gap of 100 bp
* 41392 43172: contig of 1781 bp in length
* 43173 43272: gap of 100 bp
* 43273 45886: contig of 2614 bp in length
* 45887 45986: gap of 100 bp
* 45987 49186: contig of 3200 bp in length
* 49187 49286: gap of 100 bp
* 49287 52781: contig of 3495 bp in length
* 52782 52881: gap of 100 bp
* 52882 57288: contig of 4407 bp in length
* 57289 57388: gap of 100 bp
* 57389 62317: contig of 4929 bp in length
* 62318 62417: gap of 100 bp
* 62418 65388: contig of 2971 bp in length
* 65389 65488: gap of 100 bp
* 65489 70061: contig of 4573 bp in length
* 70062 70161: gap of 100 bp
* 70162 75771: contig of 5610 bp in length
* 75772 75871: gap of 100 bp
* 75872 83423: contig of 7552 bp in length
* 83424 83523: gap of 100 bp
* 83524 89716: contig of 6193 bp in length
* 89717 89816: gap of 100 bp
* 89817 97026: contig of 7210 bp in length
* 97027 97126: gap of 100 bp
* 97127 104654: contig of 7528 bp in length
* 104655 104754: gap of 100 bp
* 104755 114540: contig of 9586 bp in length
* 114541 114440: gap of 100 bp
* 114441 126202: contig of 11762 bp in length
* 126203 126302: gap of 100 bp
* 126303 142308: contig of 15906 bp in length
* 142309 142308: gap of 100 bp
* 142309 159777: contig of 17469 bp in length
* 159778 159877: gap of 100 bp
* 159878 178832: contig of 18955 bp in length
* 178833 178932: gap of 100 bp
* 178933 180700: contig of 1768 bp in length.

```

FEATURES

```

source
  1. 180700
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="RP13-506N4"
  /clone.lib="RP13-13 Human Female BAC 481-768"
  1. 951
    /note="assembly-fragment"
  /note="assembly-fragment"
  clone_end:SP6
  vector_side:left"
  1052. 2064
    /note="assembly-fragment"
  2165. 3416
    /note="assembly-fragment"
  misc_feature
    /note="assembly-fragment"
  3517. 4816
    /note="assembly-fragment"
  misc_feature
    /note="assembly-fragment"
  4917. 6297
    /note="assembly-fragment"
  misc_feature
    /note="assembly-fragment"
  6398. 9144
    /note="assembly-fragment"
  misc_feature
    /note="assembly-fragment"
  9245. 11578
    /note="assembly-fragment"
  misc_feature
    /note="assembly-fragment"
  11679. 30905
    /note="assembly-fragment"

```

```

misc_feature 31006..32606
  /note="assembly-fragment"
misc_feature 32707..34256
  /note="assembly-fragment"
misc_feature 34357..37618
  /note="assembly-fragment"
misc_feature 37719..39131
  /note="assembly-fragment"
misc_feature 39232..41291
  /note="assembly-fragment"
misc_feature 41392..43172
  /note="assembly-fragment"
misc_feature 43273..45886
  /note="assembly-fragment"
misc_feature 45987..49186
  /note="assembly-fragment"
misc_feature 49287..52781
  /note="assembly-fragment"
misc_feature 52882..57288
  /note="assembly-fragment"
misc_feature 57389..62317
  /note="assembly-fragment"
misc_feature 62418..65388
  /note="assembly-fragment"
misc_feature 65489..70061
  /note="assembly-fragment"
misc_feature 70162..75771
  /note="assembly-fragment"
misc_feature 75872..83423
  /note="assembly-fragment"
misc_feature 83524..89716
  /note="assembly-fragment"
misc_feature 89817..97026
  /note="assembly-fragment"
misc_feature 97127..104654
  /note="assembly-fragment"
misc_feature 104755..114540
  /note="assembly-fragment"
misc_feature 114441..126202
  /note="assembly-fragment"
misc_feature 126303..142208
  /note="assembly-fragment"
misc_feature 142309..159777
  /note="assembly-fragment"

Query Match 64.0%; Score 16; DB 2; Length 180700;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 aggtgcatgcgcagca 18
DB 25268 AGGTGCATGCAGCA 25253

```

```

RESULT 37
AC007809 183439 bp DNA INV 24-FEB-2001
LOCUS Drosophila melanogaster, chromosome 3R, region 88C-88C, BAC clone
DEFINITION BAC45M03, complete sequence.
ACCESSION AC007809
VERSION AC007809.8 GI:13122709
KEYWORDS HTG;
SOURCE fruit fly;
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 183439)
Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brannon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,I.E., Doyle,C., Dresnek,D., Farfan,D.,
Fertlert,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,

```

Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacleby, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svirska, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C. 88c Sequencing of *Drosophila* chromosome 3R, region 88c-88c Unpublished

2 (bases 1 to 183439)
 Celniker, S.E., Abmayyan, A., Arcata, T.T., Baxter, E., Blazek, R.G., Butenoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomolan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleby, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Shih, E., Svirska, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zierman, L.L. and Rubin, G.M.
 Direct Submission
 Submitted (14-JUN-1999) *Drosophila* Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
 On Feb 24, 2001 this sequence version replaced gi:5623906.
 Sequence submitted by:
 Berkeley *Drosophila* Genome Project
 Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.berkeley.edu.

location/Qualifiers

1. 183439
 /organism="Drosophila melanogaster"
 /strain="y; cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="3R"
 /map="88C-88C"
 /clone="BACR45M03 (D718)"
 /clone_lib="RPCT-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
 PBace3.6)"
 BASE COUNT 53300 a 38849 c 38854 g 52436 t
 ORIGIN

Query Match 64.0%; Score 16; DB 3; Length 183439;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 10 tcgccagattcagat 25
 Db 66707 TCCGCACATTTCAT 66722

RESULT 38
 AE003705
 LOCUS 232744 bp DNA INV 05-OCT-2000
 DEFINITION *Drosophila melanogaster* genomic scaffold 142000013386035 section 30
 of 105, complete sequence.
 ACCESSION AE003705
 VERSION AE003705.1 GI:7299886
 KEYWORDS HTG.
 SOURCE *Drosophila*
 ORGANISM *Drosophila melanogaster*
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Phlebotomidae; Drosophilidae; *Drosophila*.
 1 (bases 1 to 232744)
 REFERENCE Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,
 Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,

George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazek, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, M.K., G.L., Abell, J.F., Agbayani, A., An, H.J., Andrews, P., Bannister, K., Baldwin, D., Baliga, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhattacharjee, P., Bolashkov, S., Botkova, D., Botchan, M.R., Bouck, J., Brockdorff, I., Brottier, P., Butts, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davey, L.B., Davies, P., de Pablo, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Dou, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferreira, S., Fleischmann, W., Foster, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorelli, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Helman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.D., Kraft, C., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Ku, P.D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., Miki, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleby, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, R.D., Puri, V., Reese, M.G., Rehrer, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shie, B.C., Siden-Kimms, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spindler, A.C., Stapleton, M., Strong, R., Sun, E., Svirska, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wasserman, D.A., Weinstock, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhang, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
 The genome sequence of *Drosophila* melanogaster
 Science 287 (5461), 2185-2195 (2000)
 20196006

LOCATION/Qualifiers
 1. 232744
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="3R"
 /complement(join(29274..29439,29524..30038,30107..30980))
 /gene="CG8464"

product="CT24781"
 /db_xref="FLYBASE:FBan0008464"
 /db_xref="FLYBASE:FBgn0038233"
 complement(29274..29439,29524..30038,30107..30980)
 /gene="CG8464"
 /map="88C4-88C6"
 /db_xref="FLYBASE:FBan0008464"
 /db_xref="FLYBASE:FBgn0038233"
 /protein_id="AAE5062.1"
 /db_xref="GI:7299887"
 /translation="MALRSGSHRELVIFKRCIASPVLSHANRRSSOLAIKEDPSN
 GNSFOYONCEOEKQKRWRLVRFVPSLSGAASALIOEDITPTASKMGRKRD
 ENFIADVACADSVYIEIKDRHPDPSGOGTITVNSNGSGLTENGCLTNAHVI
 NKPTHTVOVLSGRFPTATIEVDOTSDLATIRIVNNISVWRKLSSTLVKGEVAV
 ALGSPALSTVTAIGVISTORASQELGKRNINVLQDDAATPENSGLVNLNDE
 AIGVNSMKVTAGISFAIPIDYKVFLERAEKRRKGSAYKTYPVRYMGITWLTLP

* the accession number will be preserved.

FEATURES
Source
Location/Qualifiers
1..237119
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

BASE COUNT 67689 a 49909 c 50172 g 69349 t

ORIGIN

Query Match 64.0%; Score 16; DB 2; Length 237119;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tgcgcagcatctgcat 25
|||||

Db 103681 TCCGCACGATTTCCAT 103666

RESULT 40
AE003424
LOCUS
DEFINITION Drosophila melanogaster genomic scaffold 14200013386054 section 8
of 35, complete sequence.
ACCESSION AE003424 AE002566
VERSION AE003424.1 GI:7290316
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 289090)

REFERENCE
AUTHORS
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Binkley,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazer,V.G., Champagne,M., Pfeiffer,B.D.,
Man,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gaber
Mikos,G.L., Brill,J.F., Agbayani,A., An,H.J.,
Andrews-Pfankuch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Bernan,B.P., Bhandari,D., Bolintsov,S., Borokov,D.,
Botchan,M.R., Boucek,J., Brinkstein,P., Brothier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Eganellista,C.C., Ferraz,C., Fiertera,S.,
Fleischmann,W., Foster,C., Gabrielian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wel,M.H., Ibegwan,C., Jalili,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kensin,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Krall,C.,
Krivitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Matzel,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,J.J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.R., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusken,D.R., Pauley,J.M., Palazolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yen,R.F.,
Zaveri,J.S., Zhu,D., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006

REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 289090)
Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,J.C.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA

FEATURES
Source
Location/Qualifiers
1..289090
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="X"
complement(join(<9373..9459,9519..9720,9784..10373,
10559..11005,11151..11813,12630..12711,13406..13521,
13582..14921,15552..16167,16573..>16587))
/gene="EG:BACH48C10.4"
/product="CT133606"
/db_xref="FLYBASE:FBan0014047"
/db_xref="FLYBASE:FBgn0040390"
complement(<9373..>16587)
/gene="EG:BACH48C10.4"
/map="2P5-2P5"
/note="CG14047"
/db_xref="FLYBASE:FBan0014047"
/db_xref="FLYBASE:FBgn0040390"
complement(join(9373..9459,9519..9720,9784..10373,
10559..11005,11151..11813,12630..12711,13406..13521,
13582..14921,15552..16167,16573..16587))
/gene="EG:BACH48C10.4"
/note="EG:BACH48C10.4 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0014047"
/db_xref="FLYBASE:FBgn0040390"
/protein_id="AA045777.1"
/db_xref="GI:7290317"
/translation="MDSILAAVGAGGVVEAAGGPRVQVTFYTLTELFOQRTNNS
HGHSATGATGATRTADRSRGNHGHGSHVEQIELIDCKRIINTKGTPEKSAHL
ANMKKQLKQOQOARLAIARQSEVNSADELDLPNSDEQDERTYTWRQLEKRT
KELNLKENGSGARSAPNFGKRLSGVEELSMATSDIYSISPAEVTYKRTTDSVG
VAMNRSRSGEGPVTTGSRRLRDPVVSRLSPNHCAGAGGSGPRPLPPRRTRV
VMOGCEPRPRKNOVKATVITETTRIMRSPTTGTSGSAGATGSGSPAPVAA
CHRCIPEPTHAOLSPDKMEHOTCKLESPPQVTTTLEKQEDLSLMLGLAOLA
PAATLGGERTIPEENSTPIAVVPTPSPVLTKTSHVMDSSGSSNGGCTSTSTA
TTTKOPROATIENTEPEDSDSPLEEPPEPMSALRFTMSLSLEKIPEDMDA
ADLIDLDLSEKNGHDSNPQDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE
GAWTNGAGVVSXKMSFPESRAFDIKYGRNAGDAGQOHQOGRASPEPDMQECTS
GATSGEWCCTPSGSDNDQDLQINSENTSSPTKSTSLINDDTDELMDLMLA
PPMTASTINGLPRKLEPLFEETESDEKQDSDDKTKEATTNHLYEDSAGSS
SDSEVPAPPLVPVPRVREDPEASQDENLAMEDQEDHASSVERLPLNPLATRS
RPPPLTPAKAAATPIGMNATMTTKTTTMTTPSRAPSTITISSADTSSGGRAS
GPPRPVPPPPRRLLTQTLIAGQATWLLGGCRRRRORRGRNRYRVMDNATNE
YQDAGVSGPKPTTVINDNDLPAAGQOATLSHFTNSLNGKADPEPSTHPVAAG
SELPGSASTTGGCARTTCVPAAPRMKRTFRAGNAGQOPHNSGGLDLRAVRS
NDTSGMDTPTVNRPNRPSAGALGSTSNNSRSSSHNSGSLPAGGLTRAGVEP
GGQQLVVOAGGAGSAGVGGSGAGSNACSSNNKSGKGLGLETLARERYSMSGS
RSQRYRETTTSTTLMAATEMKRSKMSDSEKRSWPKRAAAASLTLRRSSYMGAG
AISTTTTNTTTEAEFLTEYGRSKRLINFKRNSSETTASSTSPNFGAIGEO
POPSHSGSLLLQKLPPGQSQODSVENAGMSPRKDNKPSLNRMLMQITKRRRTNS
VSGIYAGS"
complement(join(<18933..20031,20093..20270,20329..21208,
21574..22759,24311..24418,25054..25330,25668..26541))
/gene="EG:BACH7M4.1"
/product="CT133604"
/db_xref="FLYBASE:FBan0014045"
/db_xref="FLYBASE:FBgn0040387"
complement(<18933..>26541)
/gene="EG:BACH7M4.1"
/note="CG14045"
/map="2P5-2P5"
/db_xref="FLYBASE:FBan0014045"
/db_xref="FLYBASE:FBgn0040387"
complement(join(18945..20031,20093..20270,20329..21208,
21574..22759,24311..24418,25054..25330,25668..26541))
/gene="EG:BACH7M4.1"

CDS

gene

mRNA

CDS


```

/note="EG: BACH7M4.1 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0014045"
/db_xref="FLYBASE:FBgn0040387"
/protein_id="AAF45778.1"
/db_xref="GI:7290318"
/translation="MPTMTMRHRRSSSSAVVEESRGRRGVGVGVGADANKENRGVH
MSSPSGNSALITLQDLSNVHCKSPORRSPSGSPRATPOLALRCLEPTTGGAVL
EDSOLSSSRMDTTLDMALALIESAKREVRCKTLPGACATITSTTILANMAMSESS
VHEMEVTRPTHLKRVRRNRNPKTTTINOTOSHQAKLEPLQLVSTRCLSFSSSS
SASSDLEDOVAKRSSLASPTTPPSHCTTTTSSISSSSSSNGADMEASQSGSID
VSIADAKEOOLNINAVIRCDLQSHSGSGSINAYKVALSGAOPPGYGGHSGSGS
MSSGYOFAVHRHGRPYFDORFNQISGSEETAGYLOLAWMHIDHNSVPTQSE
MOARVKSGETERTERAKLEIEMNGENAKRSEPLFGTSPINELVHPDSGASAKH
AOACRPTSRHSQPKANAGODOKETFAADODKODSDPAEMAAATVTPPOKPVAC
SGSGSNSAMALNDEVISIDSNGKEDPLPOPOPKLSKALHQRADENLFLR
FLIEDPPADCANSTTQAOATGSASASAKANENAHNLNGISGRKQSTPMPSGS
SVGAVRQOQGRPTPTMTKRLTRTEERGFSTVWTHPPVEKIEAGLSADCGIILPG
DYVLEVKHNVVTEADVNLIRSGSSLTLEIFRRSGAGATTITSDIGONNVHIS
TRIGAAGIGSEEHILTNTAGTTTMSLORTSTRIOPLANSRPRATACSGTTSI
EAKRRLHLPOVTSKESIVPYVNRRLFLQLLSRQNTAALHREVDKPYQICGR
KDLISPDHRTLFOVNDLRLIADTLEQSSDOQDEPOMNASKVLSKTITATCA
AYKRYCNGIKRACDVLNKSRQTSIEIATTEPAVRKRPDLTMTIHRPLQHERLL
KLMOILAGNCHVDTREHKNSTVINELOAAVREITVSSGLMEPLGEGRPILLQDLES
RMVPTCKKPTLAVQGRQWILGDSLHVEGSGVPYMTLLFSOILIFAKYSRDLVFI
TEEPPIANVYDSCFHMRRKKTTERLTVDPMGRLAESPCTGCAPDLTRTPKGRKRS
LIRAPSLLEKAWONLQROIFLVNALGSTPLSPDSDVYNTLVPLSDIGLTTA
SMGSMKLPISDISHLKQOQKQVRLFKSRKRLSDSDYENLQRLASAVRHHCATTTSS
ATQTOHPOLPLCTTSLPSRTNSPARNOHRHNGGHHGSGVPGVGVSGCGCTTSS
CLOSSASLTOTQVHTOTLNHSOTHSOTQVLSLSCKCLVYIPVTSSEPNQHA
SOKLLEFSLSVSGSFIDSGGVGVGVVGLPTPTPNPNTSPTSOCPNSAFSDNE
ATSTTGKGNPILAEFGSMWLLFLDLOLHVNVLDPSTVDIVDECITFLGDEDEGNE
GCHGDESDDEDSDVYDPRGMLPTRPTPPDSDL"
/complement(<53568..>54629)
/gene="EG: BACH7M4.4"
/product="CT33234"
/db_xref="FLYBASE:FBan0012496"
/db_xref="FLYBASE:FBgn0040385"
/evidence=not_experimental
/complement(<53568..>54629)
/gene="EG: BACH7M4.4"
/note="CG12496"
/map="2P6-3A1"
/db_xref="FLYBASE:FBan0012496"
/db_xref="FLYBASE:FBgn0040385"
/evidence=not_experimental
/complement(53568..54629)
/gene="EG: BACH7M4.4"
/note="EG: BACH7M4.4 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0012496"
/db_xref="FLYBASE:FBgn0040385"
/evidence=not_experimental
/protein_id="AAF45779.1"
/db_xref="GI:7290319"
/translation="MLKWTASQRLNEIPVDDQANOPRRRLQRRORRATIANK
ENVPVGYTSPVPTIRLNSPLVLAISPLIRTPPAVYRSAPQGVSVRAVATL
PRFADYSPNGLPSGHLALGLAPLDLFLOQPIRYVSSGACVAKLRRIIDEPVAT
LEPELQETAVPEPTPKVEKVRSTPQAPASQDDTLDRLIDALLDSACKANASSKK
PRSTFNLRRTLVKQOMESNCPQVLSPPASDLSFGCLVPLPMNLMPTE
PASPLAKIPHNLLVOLATPVAFKSPACDNFTCLFTPVALKRTRKREIVAKSSPKRYK
VDERNVEFVGLAPSSIV"
/complement(<70555..71849,71925..>71976)
/gene="gt"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/product="CF23972"
/db_xref="FLYBASE:FBan0007952"
/db_xref="FLYBASE:FBgn0001150"
/complement(<70555..>71976)
/gene="gt"
/note="CG7952"
/map="3A1-3A1"

```

```

Query Match          /db_xref="FLYBASE:FBan0007952"
Best Local Similarity 64.0%; Score 16; DB 3; Length 289090;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 gtgcatcgccagcat 20
Db 13743 GTGCATCGCCAGCATT 13758

RESULT 41
AF374832/c
LOCUS      AF374832/c
DEFINITION Saxifraga fortunei 26S ribosomal RNA gene, partial sequence.
ACCESSION AF374832
VERSION    AF374832.1 GI:14210120
KEYWORDS   .
SOURCE      Saxifraga fortunei.
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Saxifragales; Saxifragaceae; Saxifraga.
REFERENCE   1 (bases 1 to 1185)
  Solits,D.E., Kuzoff,R.K., Mort,M.M., Zanis,M., Fishbein,M.,
  Huffor,L., Koontz,J. and Arroyo,M.
  Elucidating deep-level phylogenetic relationships in Saxifragaceae
  using sequences for six chloroplastic and nuclear DNA regions
  Ann. Mo. Bot. Gard. (2001) In press
  2 (bases 1 to 1185)
  Solits,D.E., Kuzoff,R.K., Mort,M.M., Zanis,M., Fishbein,M.,
  Huffor,L., Koontz,J. and Arroyo,M.
  Direct Submission
  Submitted (25-APR-2001) Department of Integrativ Biology
  Washington State University, Science Hall, Pullman, WA 99164, USA
  Location/Qualifiers
    1..1185
      /organism="Saxifraga fortunei"
      /db_xref="taxon:159961"
      <1..>1185
      /product="26S ribosomal RNA"
      BASE COUNT  279 a 273 c 381 g 252 t
      ORIGIN
        1
        2

Query Match          60.0%; Score 15; DB 8; Length 1185;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 agtgcacgcagcagc 17
Db 441 AGTGCATCGCCAGC 427

RESULT 42
NTPTLG/c
LOCUS      NTPTLG 1369 bp DNA
DEFINITION N.tabacum gene for pectate lyase.
ACCESSION X61102
VERSION    X61102.1 GI:19981
KEYWORDS   pectate lyase.
SOURCE      common tobacco.
ORGANISM    Nicotiana tabacum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
REFERENCE   1 (bases 1 to 1369)
  Lonsdale,D.M.
  Direct Submission
  Submitted (25-JUL-1991) D.M. Lonsdale, Cambridge Laboratory, Centre
  for Plant Science Research, Colney Lane, Norwich, Norfolk NR4 7UJ,
  UK
  2 (bases 1 to 1369)

```

AUTHORS Rogers, H.J., Harvey, A. and Lonsdale, D.M.
 TITLE Isolation and characterization of a tobacco gene with homology to pectate lyase which is specifically expressed during microsporogenesis
 JOURNAL Plant Mol. Biol. 20 (3), 493-502 (1992)
 MEDLINE 93043039
 COMMENT See X61101 for N. tabacum pectate lyase mRNA (partial).
 FEATURES
 source Location/Qualifiers
 1.1389
 /organism="Nicotiana tabacum"
 /strain="Samsun"
 /db_xref="taxon:4097"
 /dev_stage="mature pollen"
 /clone_1lb="lambda-EMBL3"
 /clone="G10"
 <1..577
 /number=1
 /number=1
 join(1..577,676..976,1074..1389)
 cds
 /product="pectate lyase"
 /protein_id="CAA43414.1"
 /db_xref="GI:19982"
 /db_xref="SWISS-PROT:P40972"
 /translation="MDVYRIRISVFLLVLTFFALTTATNIPRRLSNKKYKGPRA
 ENAIDKWCRCDEPMAENRQKADGALGFGSNAIGKLGRIYVTDNSDDVDPRGT
 LRYGVIOKEPLMIIFGKNMKIKLSRELLVTSNKTIDGKGFVNHIONGAGIKIOSASNI
 IISNLRHNTVPPTGGLRESDEHYGLSGSDGDISIFSSHDIWIDHISMRATDGL
 IDVAASNTITTSNCHFTDHEKVMLEFGANDHYLDKDKMTILAVNHEKRLDQRPK
 REGFPLVNDYTHMERYAIGSSGATITISQGNRFIADELLVKEVYREKLTASVAE
 WKKWTWISDGDMDENGATFTPSGDNDLIDHNLIKPEPSSKVGILTKFGALSCV
 KGRPC"
 join(<1..577,676..976,1074..>1389)
 578..675
 /number=1
 676..976
 /number=2
 977..1073
 /number=2
 1074..>1389
 /number=3
 BASE COUNT 427 a 270 c 305 g 387 t
 ORIGIN
 mRNA
 intron
 exon
 intron
 exon
 Query Match 60.0%; Score 15; DB 8; Length 1389;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 gcatcgccagcattt 21
 ||||||||||||||||
 Db 162 GCATCGCCGACGATTT 148
 RESULT 43
 NTEPCL1/c
 LOCUS NTEPCL1 1713 bp DNA
 DEFINITION N. tabacum gene for pectate lyase.
 ACCESSION X67158.1 S47561
 VERSION X67158.1 GI:19907
 KEYWORDS pectate lyase.
 SOURCE Common tobacco.
 ORGANISM Nicotiana tabacum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
 1 (bases 1 to 1713)
 Lonsdale, D.M.
 Direct Submission
 Submitted (03-JUL-1992) D.M. Lonsdale, Cambridge Laboratory, Centre
 for Plant Science Research, Colney Lane, Norwich Norfolk NR4 7UJ,
 UK
 REFERENCE 2 (bases 1 to 1713)
 JOURNAL
 TITLE
 AUTHORS
 REFERENCES

AUTHORS Rogers, H.J., Harvey, A. and Lonsdale, D.M.
 TITLE Isolation and characterization of a tobacco gene with homology to pectate lyase which is specifically expressed during microsporogenesis
 JOURNAL Plant Mol. Biol. 20 (3), 493-502 (1992)
 MEDLINE 93043039
 COMMENT
 FEATURES
 source Location/Qualifiers
 1.1713
 /organism="Nicotiana tabacum"
 /strain="Samsun"
 /db_xref="taxon:4097"
 /germline
 /dev_stage="mature pollen"
 /clone_1lb="lambda EMBL3"
 /clone="G10"
 108..1713
 /number=1
 108..901
 /number=1
 join(108..901,1000..1300,1398..1713)
 325..399
 /number=1
 join(325..901,1000..1300,1398..1713)
 cds
 /product="pectate lyase"
 /protein_id="CAA47630.1"
 /db_xref="GI:19908"
 /db_xref="SWISS-PROT:P40972"
 /translation="MDVYRIRISVFLLVLTFFALTTATNIPRRLSNKKYKGPRA
 ENAIDKWCRCDEPMAENRQKADGALGFGSNAIGKLGRIYVTDNSDDVDPRGT
 LRYGVIOKEPLMIIFGKNMKIKLSRELLVTSNKTIDGKGFVNHIONGAGIKIOSASNI
 IISNLRHNTVPPTGGLRESDEHYGLSGSDGDISIFSSHDIWIDHISMRATDGL
 IDVAASNTITTSNCHFTDHEKVMLEFGANDHYLDKDKMTITLAVNHEKRLDQRPK
 REGFPLVNDYTHMERYAIGSSGATITISQGNRFIADELLVKEVYREKLTASVAE
 WKKWTWISDGDMDENGATFTPSGDNDLIDHNLIKPEPSSKVGILTKFGALSCV
 KGRPC"
 join(400..901,1000..1300,1398..1710)
 /product="pectate lyase"
 902..999
 /number=1
 1000..1300
 /number=2
 1301..1397
 /number=2
 1398..1713
 /number=3
 BASE COUNT 546 a 338 c 353 g 476 t
 ORIGIN
 mat_peptide
 intron
 exon
 intron
 exon
 Query Match 60.0%; Score 15; DB 8; Length 1713;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 gcatcgccagcattt 21
 ||||||||||||||||
 Db 486 GCATCGCCGACGATTT 472
 RESULT 44
 HSSYNTR05/c
 LOCUS HSSYNTR05 1907 bp mRNA
 DEFINITION Homo sapiens mRNA for syntrophin 5.
 ACCESSION AJ003029
 VERSION AJ003029.1 GI:8247276
 KEYWORDS syntrophin.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 1907)
 Piuoso, G., Belisio, A., Puca, G.A. and Nigro, V.
 Identification of a novel syntrophin-like cDNA
 JOURNAL
 TITLE
 AUTHORS
 REFERENCES
 REFERENCE 2 (bases 1 to 1907)

AUTHORS Nigro, V.
 TITLE Direct Submission
 JOURNAL Submitted (11-NOV-1997) Nigro V., Istituto di Patologia Generale ed Oncologia, Seconda Università degli Studi di Napoli, Larghetto S. Aniello a Caponapoli 2, 80138, ITALY

FEATURES
 source
 1. .1907
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="p25"
 132. .1751
 /codon_start=1
 /product="syntrrophin 5"
 /protein_id="CA892969.1"
 /db_xref="GI:8247277"

CDS

translation="MGTEGPPPPASRGCGCLLPARTTTTALLYDESENAYDIR
 LKLRVLTLOKDVQVCGSHQGRNRVTLRQPGGLGLSTKGSSENNPVIISK
 IFEDQADDTGMLFVDAVLQVNGIHENATHEEVHLINAGDEVITVEYLKFAFA
 FLKPLGSGPSSDSSGASPLDFSGLIHNGNSSTAPSSPSPIAKDPRYKRMID
 TLVPLSMARISRYKACTEKIRMAAFEVLLADGVSGILRPYADGCTDMLRAVSANI
 RELTQNMKMANKCCSPSDOVYHMGVYNEKIQGDSQTRPRKFLAKGSPYFSTP
 PVSTFDWRARERTHLCVELKRVKFWLTEDCWQANDLYGLDDFDEDDRPCEFSIV
 AGHGSHVFNVEIGSELAMMEKSFORATFMEVORTGSRVYSCWQGMELCFYVDFALG
 FTCPESKTKNYLMRFKFSQLKSSDDGKTGVKLLFQWLDTRKQLEMKELFEQDLRAVLH
 CIHSFIKAVASVDPGFMDSQSLARKYVSS"

BASE COUNT 474 a 476 c 537 g 420 t
 ORIGIN

Query Match 60.0%; Score 15; DB 9; Length 1907;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgcgcagcatc 22
 |||
 Db 570 CATGCCAGCATTTTC 556

RESULT 45
 AB048915/C
 LOCUS AB048915 2616 bp mRNA PRI 20-SEP-2000
 DEFINITION Macaca fascicularis brain cDNA, clone:QnpA-17839.
 ACCESSION AB048915
 VERSION AB048915.1 GI:10241989
 KEYWORDS fis (full insert sequence).
 SOURCE Macaca fascicularis adult male brain parietal lobe cDNA to mRNA,
 clone_1lb:macaque brain cDNA library QnpA clone:QnpA-17839.
 ORGANISM Macaca fascicularis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 Cercopithecinae; Macaca.

REFERENCE
 AUTHORS 1 (sites)
 TITLE Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M.,
 Terao, K., Suzuki, Y., Sugano, S., and Hashimoto, K.
 Isolation of full-length cDNA clones from macaque brain cDNA
 libraries
 Unpublished (2000)
 2 (bases 1 to 2616)

JOURNAL
 REFERENCE Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.
 TITLE Direct Submission
 JOURNAL Submitted (14-SEP-2000) to the DDBJ/EMBL/GenBank databases.
 Katsuyuki Hashimoto, National Institute of Infectious Diseases,
 Division of Genetic Resources, 23-1 Toyama 1-chome, Shinjuku-ku,
 Tokyo 162-8640, Japan (E-mail: khashimoto@nii.ac.jp,
 URL: http://www.nii.go.jp/yoken/genbank/,
 Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)

COMMENT

Vector: pME18S-FL3 (Acc. No. AB009664)
 R. Site1: DraIII (CACTGCTG)
 R. Site2: DraIII (CACCATGTC)
 Description: 1st strand cDNA was primed with an oligo(dT) primer
 [ATGTCGCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized

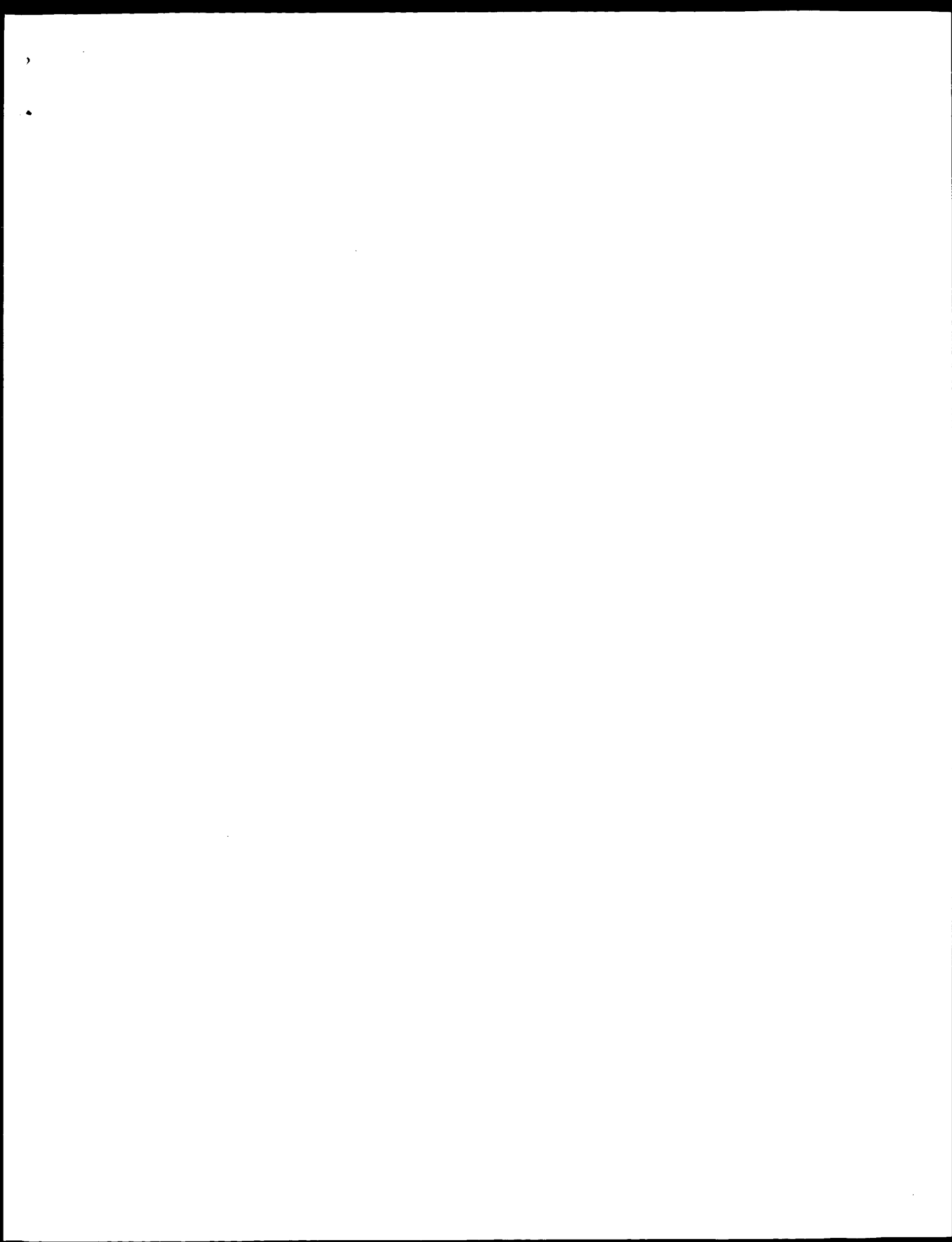
using specific 5' and 3' primers and amplified by PCR. The PCR
 product was digested with SfiI and size selection was performed to
 exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
 into distinct DraIII sites of pME18S-FL3. XhoI sites just outside
 the DraIII sites can be used to isolate the cDNA insert. Libraries
 were constructed by Sugano et al. (University of Tokyo, Institute of
 Medical Science). Custom primer used for sequencing
 (5' end primer [CTTCTGCTCTTAAGCTCCG];
 3' end primer [CGACCTGCAGCTCGACACAC]).
 Location/Qualifiers
 1. .2616
 /organism="Macaca fascicularis"
 /db_xref="taxon:9541"
 /clone_1lb="QnpA-17839"
 /dev_stage="adult"
 /sex="male"
 /tissue_type="brain parietal lobe"

BASE COUNT 739 a 527 c 572 g 778 t
 ORIGIN

Query Match 60.0%; Score 15; DB 9; Length 2616;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgcgcagcatc 22
 |||
 Db 1743 CATGCCAGCATTTTC 1729

Search completed: December 26, 2001, 12:52:19
 Job time: 8861 sec




```

85      13      52.0      1340      21      AAA93622
86      13      52.0      1357      21      AAA15971
87      13      52.0      1364      20      AAX97878
88      13      52.0      1404      20      AAV84622
89      13      52.0      1404      21      AAC48044
90      13      52.0      1419      20      AAZ12176
91      13      52.0      1419      20      AAZ12177
92      13      52.0      1419      20      AAZ12178
93      13      52.0      1430      22      AAF93795
94      13      52.0      1441      21      AAC35963
95      13      52.0      1467      22      AAH65041
96      13      52.0      1475      22      AAH99802
97      13      52.0      1481      22      AAF27856
98      13      52.0      1497      20      AAX97831
99      13      52.0      1515      20      AAX98025
100     13      52.0      1515      20      AAV84440

```

ALIGNMENTS

RESULT 1

AA162941 standard; DNA: 839 BP.

AA162941; (first entry)

Human genomic DNA SEQ ID NO 269.

Human; noctropic; neuroprotective; cytostatic; dermatological; virucide;
 immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
 antiparkinsonian; antisticking; antianemic; antiallergic; cancer;
 antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
 antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
 antiparasitic; cardiac; immune disorder; cardiovascular disorder;
 neurological disease; infection; nephrotropic; gene therapy; vaccine;
 ds.

Homo sapiens.

WO200155449-A1.

02-AUG-2001.

17-JAN-2001; 2001WO-US01346.

31-JAN-2000; 2000US-0179065.

04-FEB-2000; 2000US-0180628.

19-MAY-2000; 2000US-0205515.

07-JUL-2000; 2000US-0216880.

14-JUL-2000; 2000US-0218290.

14-AUG-2000; 2000US-0225447.

01-SEP-2000; 2000US-0230437.

06-SEP-2000; 2000US-0231243.

25-SEP-2000; 2000US-0234897.

29-SEP-2000; 2000US-0236367.

13-OCT-2000; 2000US-0239337.

08-NOV-2000; 2000US-0246476.

08-NOV-2000; 2000US-0246477.

08-NOV-2000; 2000US-0246525.

08-NOV-2000; 2000US-0246526.

08-NOV-2000; 2000US-0246528.

17-NOV-2000; 2000US-0249210.

17-NOV-2000; 2000US-0249211.

17-NOV-2000; 2000US-0249214.

01-DEC-2000; 2000US-0250160.

01-DEC-2000; 2000US-0250391.

05-DEC-2000; 2000US-0251030.

05-DEC-2000; 2000US-0251988.

Human peroxisome a
 Human protein clon
 Human secreted pro
 Human secreted pro
 Zee mays DNA fragm
 Neisseria meningit
 Neisseria gonorrhoe
 Human cDNA encodin
 Arabidopsis thaliana
 C glutamicum codin
 Human protein enco
 Human ND6 cDNA.
 Human secreted pro
 Human secreted pro
 Human secreted pro

05-DEC-2000; 2000US-0256719.
 06-DEC-2000; 2000US-0251479.
 08-DEC-2000; 2000US-0251989.
 08-DEC-2000; 2000US-0251990.
 11-DEC-2000; 2000US-0254097.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPL: 2001-476225/51.

Novel plasma membrane associated proteins useful for diagnosing,
 treating, preventing and/or prognosing disorders related to the
 proteins, including cancer, immune response and neuronal disorders

Example 2: SEQ ID NO 269; 532pp + Sequence listing; English.

The invention relates to novel genes (AA162752-AA162961) and proteins
 (AAM42347-AA42415) useful for preventing, treating or ameliorating
 medical conditions e.g. by protein or gene therapy. The genes are
 isolated from a range of human tissues disclosed in the specification.
 The nucleic acids, proteins, antibodies and (ant)agonists are useful
 in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 and ovarian cancer and other cancers of the adrenal gland, bone, bone
 marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 colitis; (c) cardiovascular disorders such as myocardial ischemias;
 (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 and parasitic infections.
 Note: The sequence data for this patent did not form part of the
 printed specification, but was obtained in electronic format directly
 from WIPO at http://wipo.int/pub/published_pcl_sequences.

Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other;

Query Match 100.0%; Score 25; DB 22; Length 839;
 Best Local Similarity 100.0%; Pred. No. 3.3e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcacgcgcacatcgcacat 25
 Db 116 gcaggtgcacgcgcacatcgcacat 140

RESULT 2

AA162941 standard; DNA: 1041 BP.

12-APR-1999 (first entry)

E. coli biotin synthetase (BioB) coding sequence.

DAP aminotransferase; diaminopelargonic acid; transgenic plant;

biotin synthetase; biotin production; vitamin H; BioB; ss.

Escherichia coli.

US5869719-A.

09-FEB-1999.

30-APR-1997; 97US-0846338.

30-APR-1997; 97US-0846338.

08-MAR-1995; 95US-0401068.

PA (NOVS) NOVARTIS FINANCE CORP.
 XX
 PI Patton DA:
 XX
 DR WPI: 1999-152902/13.
 DR P-PSDB: AAM73906.
 XX
 PT Transgenic plants with high biotin levels - transformed with DNA
 PT encoding di:amino-pelargonic acid amino-transferase or biotin
 PT synthase
 PS Example 2: Column 37-40; 34pp: English.
 XX
 CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.
 CC
 SO Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 3.3e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcaggtgcatcgccagcattcgat 25
 |||||||||||||||||||||
 DB 84 gcaggtgcatcgccagcattcgat 108

RESULT 3
 AAN91329
 ID AAN91329 standard; DNA: 1084 BP.
 XX
 AC AAN91329;
 XX
 DT 15-FEB-1990 (first entry)
 XX
 DE E.coli Bio B gene.
 XX
 KW E.coli; Bio B gene; biotin.
 XX
 OS Escherichia coli.
 XX
 FT Key Location/Qualifiers
 FT CDS 24..1064
 FT /*tag=a
 XX
 PN GB2216530-A.
 XX
 PD 11-OCT-1989.
 XX
 PE 17-MAR-1989; 89GB-0006210.
 XX
 PR 22-MAR-1988; 88GB-0006804.
 PR 17-MAR-1989; 89GB-0006210.
 XX
 PA (UKAG-) UK MIN. AGRIC. FISH.
 XX
 PI Pearson BM, McKee RA;
 XX
 DR WPI: 1989-295085/41. P-PSDB P91392
 XX
 PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.
 PS Table 3; page 33-4; 52pp: English.
 XX
 CC The gene can be used in a plasmid for expression of enzymes of the biotin

CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pMA91, pMA36C, PKV49 and pCK499, and plasmid pCK965 for
 CC Lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.
 CC
 SO Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 3.3e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcaggtgcatcgccagcattcgat 25
 |||||||||||||||||||||
 DB 107 gcaggtgcatcgccagcattcgat 131

RESULT 4
 AAN60496
 ID AAN60496 standard; DNA: 1121 BP.
 XX
 AC AAN60496;
 XX
 DT 17-OCT-1991 (first entry)
 XX
 DE Sequence encoding biotin synthesising enzyme.
 XX
 KW Biotin synthetic enzyme; E.coli; deshydrobiotin; ds.
 XX
 FH Key Location/Qualifiers
 FT CDS 42..1082
 FT /*tag=a
 XX
 PN JP61149091-A.
 XX
 PD 07-JUL-1986.
 XX
 PE 24-DEC-1984; 84JP-0272605.
 XX
 PR 24-DEC-1984; 84JP-0272605.
 XX
 PA (NIPS) NIPPON SODA KK.
 XX
 DR WPI: 1986-216622/33.
 DR P-PSDB: AAP60536.
 XX
 PT Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
 XX
 PS Disclosure: Page 534; 23pp: Japanese.
 XX
 CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing deshydrobiotin.
 CC
 SO Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 3.3e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcaggtgcatcgccagcattcgat 25
 |||||||||||||||||||||
 DB 125 gcaggtgcatcgccagcattcgat 149

RESULT 5
 AA062386
 ID AA062386 standard; DNA: 5872 BP.
 XX

AC	AA062386;
XX	
DT	16-NOV-1994 (first entry)
XX	
DE	Biotin-biosynthesis genes contg. plasmid PB030A-15/9.
XX	
KW	Biotin; expression; enterobacteria; vitamin H; synthesis;
KW	plasmid; PB030A-15/9; biob; biof; bioc; biod; bioA;
KW	promoter plac; biotin synthase; KAPA synthase;
KW	8-amino-7-oxonanoate synthase; pimeloyl-CoA; PNB synthase;
KW	dethiodiolin synthase; DAA synthase;
KW	S-adenosyl-L-methionine; 8-amino-7-oxonanoate aminotransferase;
XX	seborrhoea; dermatitis; ds.
OS	Escherichia coli DSM498.
XX	
FH	Key
FT	Location/Qualifiers
FT	1..96
FT	/tag= a
FT	/function= "promoter plac"
FT	/evidence= EXPERIMENTAL
FT	23..28
FT	/tag= b
FT	/standard_name= "promoter plac"
FT	45..50
FT	/tag= c
FT	/evidence= EXPERIMENTAL
FT	/standard_name= "promoter plac"
FT	105..109
FT	/tag= d
FT	/evidence= EXPERIMENTAL
FT	/standard_name= "biob RBS no. 9"
FT	117..1157
FT	/tag= e
FT	/product= "biotin synthase"
FT	/evidence= EXPERIMENTAL
FT	/gene= "biob"
FT	/number= 1
FT	1141..1146
FT	/tag= f
FT	/standard_name= "biof RBS"
FT	1154..2311
FT	/tag= g
FT	/EC_number= 2.3.1.47
FT	/product= "KAPA synthase"
FT	/evidence= EXPERIMENTAL
FT	/gene= "biof"
FT	/number= 2
FT	/standard_name= "8-amino-7-oxonanoate synthase"
FT	2284..2288
FT	/tag= h
FT	/standard_name= "bioc RBS"
FT	2295..3050
FT	/tag= i
FT	/function= "involved in pimeloyl-CoA synthesis"
FT	/product= "protein"
FT	/gene= "bioc"
FT	/number= 3
FT	3030..3033
FT	/tag= j
FT	/standard_name= "biod RBS"
FT	3043..3753
FT	/tag= k
FT	/EC_number= 6.3.3.3
FT	/product= "PTB synthase"
FT	/evidence= EXPERIMENTAL
FT	/gene= "biod15"
FT	/number= 4
FT	/standard_name= "dethiodiolin synthase"
FT	3712..3750
FT	/tag= l
FT	/note= "biod15 substitution"
FT	3742..3746
RBS	

Query Match	100.0%; Score 25; DB 15; Length 5872;
Best Local Similarity	100.0%; Pred. No. 3,2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
1 gcaagtgatgcgcagcattcgat 25	
Db 200 gcaagtgatgcgcagcattcgat 224	
Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;	
Query Match	100.0%; Score 25; DB 15; Length 5872;
Best Local Similarity	100.0%; Pred. No. 3,2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
1 gcaagtgatgcgcagcattcgat 25	
Db 200 gcaagtgatgcgcagcattcgat 224	
Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;	


```

XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment seq ID NO: 67727.
XX
XX Hybridisation assay: genetic mapping: gene expression control:
XX Protein identification: signal transduction pathway:
XX Metabolic pathway: promoter: termination sequence: ss.
XX
OS Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0125548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149932.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.

```

PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156568.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 64.0%; Score 16; DB 21; Length 1498;
Best Local Similarity 100.0%; Pred. No. 3.5;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 aggtgcacgcacga 18
|||||
Db 84 aggtgcacgcacga 99

RESULT 7
AAC36258
ID AAC36258 standard; DNA; 1500 BP.

AC AAC36258;
XX

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 13129.

KW Hybridisation assay; genetic mapping; gene expression control;
protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
OS
XX EP1033405-A2.
PN
XX 06-SEP-2000.
ED
XX
XX
PF 25-FEB-2000; 2000EP-0301439.
PF
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139499.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139809.
PR 23-JUN-1999; 99US-0140354.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.

```
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147034.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148564.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
```

```
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 64.0%; Score 16; DB 21; Length 1500;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 aggtgcatgcgcagca 18
|||||
Db 84 aggtgcatgcgcagca 99

RESULT 8
AAH18667/c
ID AAH18667 standard; cDNA; 4674 BP.
XX
AC AAH18667;
XX
XX 26-JUN-2001 (first entry)
DE Human cDNA sequence SEQ ID NO:18909.
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
OS
XX EP1074617-A2.
PN
XX 07-FEB-2001.
PD
XX 28-JUL-2000; 2000EP-0116126.
PF

XX 29-JUL-1999: 99JP-0248036.
 PR 27-AUG-1999: 99JP-0300253.
 PR 11-JAN-2000: 2000JP-0118776.
 PR 02-MAY-2000: 2000JP-0183767.
 PR 09-JUN-2000: 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INSR.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS
 PS Claim 8: SEQ ID 18909; 2537bp + CD ROM; English.
 XX
 XX The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95883 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SO Sequence 4674 BP; 991 A; 1342 C; 1239 G; 1102 T; 0 other;

Query Match 64.0%; Score 16; DB 22; Length 4674;
 Best Local Similarity 100.0%; Pred No. 3.5;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 ggtgcatcgccagcat 19
 |||
 Db 2446 GGTGCAATCGCCAGCAT 2431

RESULT 9
 AAC93313
 ID AAC93313 standard; cDNA: 2597 BP.
 XX
 AC AAC93313:
 XX
 DT 16-FEB-2001 (first entry)
 XX
 DE Human secreted protein cDNA sequence #4.
 XX
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiinfect;
 KW vulterary; anticonvulsant; antibacterial; antifungal; aniparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein; ss.
 XX
 OS Homo sapiens.

XX
 PN W0200058495-A1.
 XX
 PD 05-OCT-2000.
 XX
 PF 23-MAR-2000; 2000WO-US07661.
 XX
 PR 26-MAR-1999: 99US-0126504.
 PR 07-JAN-2000: 2000US-0174847.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM, Komatsoulis G;
 DR WPI: 2000-611720/58.
 DR P-PSDB: AAB51383.
 XX
 PT New nucleic acid molecules encoding 45 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 PS
 PS Claim 1; Pages 327-328; 410pp; English.
 XX
 XX The invention relate to the isolation of genes AAC93310-C93354 encoding
 CC 45 human secreted proteins AAB51380-B51423. The genes can be used to
 CC generate fusion proteins by linking to the gene for the human
 CC immunoglobulin G Fc portion (AAC93301) for increasing the stability of
 CC the fusion protein as compared to the human protein only. The genes and
 CC proteins are useful for preventing, ameliorating or treating medical
 CC conditions, e.g. by protein or gene therapy. The genes are isolated
 CC from a range of human tissues disclosed in the specification. The
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
 CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 XX
 SO Sequence 2597 BP; 721 A; 604 C; 534 G; 737 T; 1 other;

Query Match 60.0%; Score 15; DB 21; Length 2597;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcaagtgcatcgcca 15
 |||
 Db 2256 gcaagtgcatcgcca 2270

RESULT 10
 AA161089/C
 ID AA161089 standard; cDNA: 5662 BP.
 XX
 AC AA161089:
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide seq ID NO 5078.
 XX
 KW Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.

XX WO200153312-A1.
 XX 26-JUL-2001.
 XX 26-DEC-2000; 2000WO-US34263.
 XX 21-JAN-2000; 2000US-0488725.
 XX 25-APR-2000; 2000US-0552317.
 XX 09-JUL-2000; 2000US-0598042.
 XX 19-JUL-2000; 2000US-0620312.
 XX 03-AUG-2000; 2000US-0653450.
 XX 14-SEP-2000; 2000US-0662191.
 XX 19-OCT-2000; 2000US-0693036.
 XX 29-NOV-2000; 2000US-0727344.
 XX (HSE-) HXSEQ INC.
 XX Tang YN, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX WPI: 2001-442253/47.
 XX P-PSDB: AAM41933.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 XX such as central nervous system injuries -
 XX Claim 1: SEQ ID NO 5078; 10078pp; English.
 XX The invention relates to human nucleic acids (AA157798-AA161369) and
 XX the encoded polypeptides (AAM38642-AAM42213) with neurotrophic,
 XX immunosuppressant and cytostatic activity. The polynucleotides are useful
 XX in gene therapy. A composition containing a polypeptide or polynucleotide
 XX of the invention may be used to treat diseases of the peripheral nervous
 XX system, such as peripheral nervous injuries, peripheral neuropathy and
 XX localized neuropathies and central nervous system diseases, such as
 XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 XX utilisation of the activities such as: Immune system suppression,
 XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 XX assays for receptor activity, arthritis and inflammation, leukaemias and
 XX C.N.S disorders.
 XX Note: The sequence data for this patent did not form part of the printed
 XX specification.
 XX Sequence 5662 BP; 1631 A; 1126 C; 1247 G; 1658 T; 0 other;
 XX
 XX Query Match 60.0%; Score 15; DB 22; Length 5662;
 XX Best local Similarity 100.0%; Pred. No. 13;
 XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX QY 8 catgcgcagatctc 22
 XX ||||||||||||||||
 XX Db 4772 CATGCCAGCATTTTC 4758
 XX
 XX RESULT 11
 XX AAC76014/c
 XX ID AAC76014 standard; cDNA: 5773 BP.
 XX
 XX AAC76014:
 XX
 XX 08-FEB-2001 (first entry)
 XX
 XX Human OREX ORF1569 polynucleotide sequence SEQ ID NO:3137.
 XX
 XX Human: open reading frame; OREX: detection: cytostatic; hepatotropic;
 XX vulnery; antiparkinsonian; neurotrophic; neuroprotective;
 XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;

KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX
 XX Homo sapiens.
 XX
 XX WO200058473-A2.
 XX 05-OCT-2000.
 XX 31-MAR-2000; 2000WO-US08621.
 XX 31-MAR-1999; 99US-0127607.
 XX 02-APR-1999; 99US-0127636.
 XX 05-APR-1999; 99US-0127728.
 XX 30-MAR-2000; 2000US-0540763.
 XX (CURA-) CURAGEN CORP.
 XX Shinkets RA, Leach M;
 XX WPI: 2000-602362/57.
 XX P-PSDB: AAB41805.
 XX Novel nucleic acids and peptides derived from open reading frame X,
 XX useful for treating e.g. cancers, proliferative disorders,
 XX neurodegenerative disorders and cardiovascular disease -
 XX Claim 5: Page 2354-2357; 5507pp; English.
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 XX which represent the human OREX open reading frames 1 to 3161. The OREX
 XX sequences have activities such as: cytostatic; hepatotropic; vulnery;
 XX antiparkinsonian; neurotrophic; neuroprotective;
 XX osteopathic; anticonvulsant; antiarthritic; immunosuppressive;
 XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 XX antidiabetic; hypotensive; dermatological; immunosuppressive;
 XX antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 XX antithyroid; and antianaemic. The sequences can be used for determining
 XX the presence of or predisposition to, or preventing or treating
 XX pathological conditions associated with an OREX-associated disorder. The
 XX nucleic acids can be used to express OREX proteins in gene therapy
 XX vectors. The proteins and nucleic acids may be used to treat cancers,
 XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
 XX graft vs host disease, cardiovascular disease, diabetes mellitus,
 XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 XX nocturnal haemoglobinuria, antinflammatory disease; to enhance
 XX coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 XX Sequence 5773 BP; 1691 A; 1125 C; 1269 G; 1687 T; 1 other;
 XX
 XX Query Match 60.0%; Score 15; DB 21; Length 5773;
 XX Best local Similarity 100.0%; Pred. No. 12;
 XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX QY 8 catgcgcagatctc 22
 XX ||||||||||||||||
 XX Db 4886 CATGCCAGCATTTTC 4872
 XX
 XX RESULT 12
 XX AA159303/c

ID AAI59303 standard; cDNA; 5876 BP.
 AC AAI59303;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 1506.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR P-PSDB; AAM40147.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1: SEQ ID NO 1506; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 5876 BP; 1704 A; 1165 C; 1288 G; 1719 T; 0 other;

Query Match 60.0%; Score 15; DB 22; Length 5876;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgccacagatttc 22
 DB 5014 CATGCCACAGATTTC 5000

RESULT 13
 AAC93854/c
 ID AAC93854 standard; cDNA; 549 BP.
 XX
 AC AAC93854;
 XX
 DT 19-FEB-2001 (first entry)
 XX
 DE Cat flea hindgut and Malpighian tubule (HMT) cDNA, SEQ ID NO:349.
 XX
 KW Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;
 KW flea infestation; vaccine; antiparasitic; therapeutic target;
 KW diagnosis; detection; ss.
 XX
 OS Ctenocephalides felis.
 XX
 PN WO200061621-A2.
 XX
 PD 19-OCT-2000.
 XX
 PF 07-APR-2000; 2000WO-US09437.
 XX
 PR 09-APR-1999; 99US-0128704.
 XX
 PA (HESK-) HESKA CORP.
 XX
 PI Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;
 XX
 DR WPI: 2000-656323/63.
 XX
 PT Flea Malpighian tubule and head and nerve cord tissue derived nucleic
 PT acids useful for the prevention, diagnosis and treatment of flea
 PT infestations -
 XX
 PS Claim 26; Page 361; 964pp; English.
 XX
 CC The invention relates to novel cat flea (Ctenocephalides felis) nucleic
 CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
 CC or head and nerve cord (HNC) tissue. The invention also relates to the
 CC encoded proteins. The invention additionally encompasses expression
 CC constructs, recombinant viruses and recombinant cells comprising the
 CC nucleic acids of the invention, recombinant production of the proteins,
 CC antibodies against the proteins, a method of identifying inhibitors of
 CC the proteins, and compositions comprising the inhibitors for
 CC administration to an animal. The nucleic acids, and the proteins they
 CC encode may be used in the prevention, treatment and diagnosis of diseases
 CC associated with flea infestations. For example, the nucleic acids may be
 CC used to produce an HMT or HNC protein according to standard recombinant
 CC DNA methodology by inserting the nucleic acids into a host cell and
 CC culturing the cell to express the protein. The HMT and HNC nucleic acids
 CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
 CC sequences in samples. They may also be used to study the expression and
 CC function of the proteins and their role in metabolism. The HMT and HNC
 CC proteins may be used as antigens in the production of specific
 CC antibodies, and in assays to identify modulators (agonists and
 CC antagonists) of HMT and/or HNC protein expression and activity. The
 CC anti-HMT/HNC protein antibodies and antagonists may also be used to
 CC downregulate protein expression and activity. The antibodies may also be
 CC used as diagnostic agents for detecting the presence of flea polypeptides
 CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
 CC present sequence represents a cat flea HMT cDNA of the invention.
 XX
 SQ Sequence 549 BP; 155 A; 102 C; 130 G; 160 T; 2 other;

Query Match 56.0%; Score 14; DB 21; Length 549;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgccacagatttc 21

Db 421 CATGCCAGCATTT 408

RESULT 14

AAH84601/c
ID AAH84601 standard; DNA; 597 BP.

XX AC
XX AAH84601;

DT 26-SEP-2001 (first entry)

XX E. coli growth and proliferation related coding sequence SEQ ID NO:229.

DE Escherichia coli; growth; proliferation; microbial; antimicrobial;

KW bacterial infection; microorganism; ds.

XX Escherichia coli.

OS WO200134810-A2.

PN 17-MAY-2001.

XX 09-NOV-2000; 2000WO-US30950.

PR 09-NOV-1999; 99US-0164415.

XX (ELITRA) ELITRA PHARM INC.

PI Forsyth RA, Ohlsen K, Zyskind J;

XX WPI: 2001-335933/35.

DR P-PSDB: AAC98930.

XX Novel nucleic acids that inhibit Escherichia coli proliferation, useful

PT for screening for homologous genes and for designing expression vectors

XX Claim 9; Page 275-276; 522pp; English.

XX AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
CC related DNA sequences (1). AAH84500 to AAH84670 encode the E. coli
CC growth and proliferation related proteins given in AAC99078 and AAC98830
CC to AAC98999. (1) can be used as potential targets for the generation of
CC new antimicrobial agents, and for identification of compounds which
CC interact with the gene products of (1). In addition the expression of
CC (1) and the purification of the proteins, the purified proteins can be
CC used to generate reagents and screen small molecule libraries or other
CC candidate compound libraries for compounds that can be further developed
CC to yield novel antimicrobial compounds. In addition, nucleic acid probes
CC complementary to (1) that are specific for particular species of
CC microorganisms can be used to identify particular microorganism species
CC in clinical specimens, therefore, providing a rapid and dependable
CC method by which to identify the causative agents of a bacterial
CC infection. Also, antibodies generated against proteins translated from
CC mRNA transcribed from proliferation-related sequences can also be used
CC to screen for specific microorganisms that produce such proteins in a
CC species-specific manner. AAH84371 and AAH84670 represent sequencing
CC primers used in the isolation of E. coli growth and proliferation
CC related sequence, which are used in an example from the present
CC invention.

XX Sequence 597 BP; 130 A; 158 C; 184 G; 125 T; 0 other;

Query Match

Best Local Similarity 56.0%; Score 14; DB 22; Length 597;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcattgccagcat 19
Db 275 TGCATGCCAGCAT 262

RESULT 15
AAS02190/c
ID AAS02190 standard; cDNA; 668 BP.

XX AC
XX AAS02190;

DT 29-AUG-2001 (first entry)

XX DNA encoding propeptide of conopeptide L5V, K15A.

KW Gamma carboxyglutamate; conopeptide; neurological disorder; epilepsy;

KW hypoxia; anoxia; ischemia; stroke; brain; spinal cord; suffocation;

KW myocardial infarct; trauma; drowning; perinatal asphyxia; hypoglycaemia;

KW neurodegeneration; Alzheimer's disease; Huntington's disease;

KW senile dementia; Amyotrophic lateral Sclerosis; multiple sclerosis;

KW Parkinson's disease; Down's Syndrome; Korsakoff's disease; schizophrenia;

KW AIDS; acquired immunodeficiency syndrome; HIV; neuronal damage; pain;

KW seizure; chemical toxicity; addiction; dystonia; psychiatric disorder;

KW mood disorder; memory; ophthalmic; parasitic worm; ss.

XX Conus geographus.

XX Key location/Qualifiers

XX CDS 93..395

XX /*tag= a

XX /product= "Propeptide of Conopeptide L5V, K15A"

XX (UTAH) UNIV UTAH RES FOUND.

XX (COGN-) COGNETIX INC.

XX WPI: 2001-273379/28.

DR P-PSDB: AAU01503.

XX New isolated gamma-carboxyglutamate containing peptide for treating or

PT preventing neurological and psychiatric disorders e.g. epilepsy,

PT Alzheimer's disease, migraine, chemical toxicity, dystonia, anxiety,

PT and depression

XX Claim 4; Page 30-31; 102pp; English.

XX The sequence represents the coding sequence of the propeptide of gamma
CC carboxyglutamate-containing conopeptide L5V, K15A. The conopeptide is used
CC for treating or preventing disorders in which the pathophysiology
CC involves excess excitation of nerve cells by excitatory amino acids or
CC agonists of heterogeneous ionotropic glutamate receptors or heterogenous
CC G protein coupled glutamate receptors. The disorders may be neurological
CC disorders, such as: (i) seizure associated with epilepsy; (ii) a
CC neurotoxic injury associated with hypoxia, anoxia, ischemia, stroke,
CC cerebrovascular accident, brain or spinal cord trauma, myocardial
CC infarct, physical trauma, drownings, suffocation, perinatal asphyxia, or
CC hypoglycaemic events; (iii) neurodegeneration associated with Alzheimer's
CC disease, Huntington's disease, senile dementia, Amyotrophic lateral
CC sclerosis, multiple sclerosis, Parkinson's disease, Down's Syndrome,
CC Korsakoff's disease, schizophrenia, AIDS (acquired immunodeficiency
CC syndrome) dementia from HIV infection, HIV infection, multi-infarct
CC dementia, Binswanger dementia and neuronal damage associated with
CC uncontrolled seizures; (iv) pain which is a migraine, acute pain, or
CC persistent pain; (v) chemical toxicity which is addiction, morphine,
CC opiate, opioid and barbiturate tolerance; and (vi) dystonia, urinary
CC incontinence, muscle relaxation or sleep disorder. The disorders may be

CC psychiatric disorders, such as, anxiety, major depression, manic-
CC depressive illness, obsessive-compulsive disorder, schizophrenia, or mood
CC disorders (bipolar disorder, unipolar depression, dysthymia, or seasonal
CC affective disorder). The conopeptide is also used to treat memory or
CC cognitive deficits, ophthalmic indications, or to control nematodes or
CC parasitic worms.
XX
SQ Sequence 668 BP; 187 A; 160 C; 145 G; 176 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 668;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cagtcgacgcga 15
|||||
DB 101 CAGGTCGATCGCA 88

RESULT 16
AAH65704
ID AAH65704 standard; DNA; 702 BP.
XX

AC AAH65704:
XX
DT 26-SEP-2001 (first entry)
XX

DE C glutamicum coding sequence fragment SEQ ID NO: 739.
XX

KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis; ds.
XX

OS Corynebacterium glutamicum.
XX

PN EPI108790-A2.
XX

PD 20-JUN-2001.
XX

PF 18-DEC-2000; 2000EP-0127688.
XX

PR 16-DEC-1999; 99JP-0377484.
XX

PR 07-APR-2000; 2000JP-0159162.
XX

PR 03-AUG-2000; 2000JP-0280988.
XX

PA (KYO) KYOMA HAKKO KOGYO KK.
XX

PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX

DR WPI: 2001-376931/40.
XX

DR P-PSDB; AAG90485.
XX

XX
XX

PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
XX

PT mutation point of a gene, measuring expression of a gene, analysing
XX

PT expression profile or pattern of a gene and identifying homologous gene
XX

XX
XX

PS Claim 8; SEQ ID NO: 739; 246bp + Sequence Listing; English.
XX

XX
XX

CC The present invention provides a number of nucleotide and protein
XX

CC sequences from the Corynebacterium glutamicum. These
XX

CC are useful for identifying the mutation point of a gene derived from a
XX

CC mutant of corynebacterium glutamicum, measuring expression amount and
XX

CC analysing the expression profile or expression pattern of a gene derived
XX

CC from Corynebacterium glutamicum, and identifying a homologue of a gene derived
XX

CC from corynebacterium glutamicum. Corynebacterium bacteria are useful for producing
XX

CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX

CC particularly L-lysine. The present sequence is a nucleic acid described
XX

CC in the exemplification of the invention.
XX

CC Note: The sequence data for this patent did not form part of the printed
XX

CC specification, but was obtained in electronic format directly from the
XX

CC European Patent Office.
XX

SQ Sequence 702 BP; 188 A; 240 C; 156 G; 118 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 702;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 gtgcacgcgcga 18
|||||
DB 102 gtgcacgcgcga 115

RESULT 17
AAV20504/C
ID AAV20504 standard; DNA; 719 BP.
XX

AC AAV20504:
XX

DT 30-JUN-1998 (first entry)
XX

DE Conus geographus conantokin DNA.
XX

KW Conantokin; predatory cone snail; treatment; neurologic disorder;
XX

KW psychiatric disorder; anticonvulsant; neuroprotective; analgesic;
XX

KW HIV infection; ophthalmic indication; memory; learning defect;
XX

KW cognitive defect; ss.
XX

OS Conus geographus.
XX

PN Key Location/Qualifiers
XX

FT CDS 110..412
XX

FT /*tag= a
XX

FT /product= conantokin
XX

PN WC9803541-A1.
XX

PD 29-JAN-1998.
XX

PF 21-JUL-1997; 97WO-US12618.
XX

PR 22-JUL-1996; 96US-0684742.
XX

PA (COGN-) COGNETIX INC.
XX

PA (UTAH) UNIV UTAH RES FOUND.
XX

PI Abogadie FC, Colledge C, Cruz LJ, Hillyard DR, Jimenez E;
XX

PI Layer RT, McCabe RT, Olivera BM, Rivier JE, Shen GS;
XX

PI Walker C, Zhou L;
XX

DR WPI: 1998-120694/11.
XX

DR P-PSDB; AAW48210.
XX

XX
XX

PT New conantokin peptide(s) - useful for e.g. treating neurologic or
XX

PT psychiatric disorders, or the management of pain
XX

XX
XX

PS Claim 19; Pages 79-80; 122pp; English.
XX

XX
XX

CC The present sequence encodes Conus geographus conantokin, peptide
XX

CC derivatives of which can be used to treat neurologic and
XX

CC psychiatric disorders, e.g. as an anticonvulsant, neuroprotective
XX

CC or analgesic agent. Neurologic and psychiatric disorders include
XX

CC epilepsy, convulsions, neurotoxic injury (associated with
XX

CC conditions of hypoxia, anoxia or ischaemia, which typically follow
XX

CC stroke, cerebrovascular accident, brain or spinal cord trauma,
XX

CC myocardial infarct, physical trauma, drowning, suffocation,
XX

CC perinatal asphyxia or hypoglycaemic events), neurodegeneration
XX

CC (associated with Alzheimer's disease, senile dementia, Amyotrophic
XX

CC Lateral Sclerosis, Multiple Sclerosis, Parkinson's disease,
XX

CC Huntington's disease, Down's Syndrome, Korsakoff's disease,
XX

CC schizophrenia, AIDS dementia, multi-infarct dementia, Binswanger
XX

CC dementia and neuronal damage associated with uncontrolled
XX

CC seizures), chemical toxicity (such as addiction, and morphine,
XX

CC opiate), opioid and barbiturate tolerance), pain (acute, chronic,
XX

CC migraine), anxiety, major depression, manic-depressive illness,
 CC obsessive-compulsive disorder, schizophrenia and mood disorders
 CC (such as bipolar disorder, unipolar depression, dysthymia and
 CC seasonal affective disorder) and dystonia (movement disorder),
 CC sleep disorder, muscle relaxation and urinary incontinence. The
 CC peptide can also be used to treat HIV infection, ophthalmic
 CC indication and memory, learning or cognitive defects.
 CC
 SQ Sequence 719 BP; 217 A; 171 C; 149 G; 182 T; 0 other;
 XX
 XX
 Query Match 56.0%; Score 14; DB 19; Length 719;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 2 caggtgcatgcgcca 15
 |||||||||||||
 DB 118 CAGGTGCATGCCTCA 105
 RESULT 18
 AAV17130/C
 ID AAV17130 standard; DNA: 719 BP.
 XX
 AC AAV17130;
 XX
 DT 30-JUN-1998 (first entry)
 XX
 DE Conus geographus conantokin DNA.
 XX
 KW Conantokin: predatory cone snail; treatment; neurologic disorder;
 KW psychiatric disorder; anticonvulsant; neuroprotective; analgesic;
 KW HIV infection; ophthalmic indication; memory; learning defect;
 KW cognitive defect; ss.
 XX
 OS Conus geographus.
 XX
 FH Key Location/Qualifiers
 FT CDS 110..412
 FT /*tag= a
 FT /product= conantokin
 XX
 PN W09803189-A1.
 XX
 PD 29-JAN-1998.
 XX
 PE 21-JUL-1997; 97WO-US12652.
 XX
 PR 06-DEC-1996; 96US-0762377.
 PR 22-JUL-1996; 96US-0684750.
 XX
 PA (COGN-) COGNETIX INC.
 XX
 PI Layer RT, McCabe RT, Zhou L;
 XX
 DR WPI: 1998-120469/11.
 XX
 P-PSDB; AAM49989.
 XX
 PT Use of conantokin peptide(s) - for treating disorders involving
 PT excessive excitation of nerve cells by excitatory amino acids or
 PT agonists of the N-methyl-D-aspartate receptor
 XX
 PS Example 4; Pages 79-80; 122pp; English.
 XX
 CC The present sequence encodes Conus geographus conantokin, peptide
 CC derivatives of which can be used to treat neurologic and
 CC psychiatric disorders, e.g. as an anticonvulsant, neuroprotective
 CC or analgesic agent. Neurologic and psychiatric disorders include
 CC epilepsy, convulsions, neurotoxic injury (associated with
 CC conditions of hypoxia, anoxia or ischemia, which typically follow
 CC stroke, cerebrovascular accident, brain or spinal cord trauma,
 CC myocardial infarct, physical trauma, drowning, suffocation,
 CC perinatal asphyxia or hypoglycaemic events), neurodegeneration

CC (associated with Alzheimer's disease, senile dementia, Amyotrophic
 CC Lateral Sclerosis, Multiple Sclerosis, Parkinson's disease,
 CC Huntington's disease, Down's Syndrome, Korsakoff's disease,
 CC schizophrenia, AIDS dementia, multi-infarct dementia, Binswanger
 CC dementia and neuronal damage associated with uncontrolled
 CC seizures), chemical toxicity (such as addiction, and morphine,
 CC opiates, opioid and barbiturate tolerance), pain (acute, chronic,
 CC migraine), anxiety, major depression, manic-depressive illness,
 CC obsessive-compulsive disorder, schizophrenia and mood disorders
 CC (such as bipolar disorder) and dystonia (movement disorder).
 CC seasonal affective disorder) and urinary incontinence. The
 CC sleep disorder, muscle relaxation and urinary incontinence. The
 CC peptide can also be used to treat HIV infection, ophthalmic
 CC indication and memory, learning or cognitive defects.
 CC
 SQ Sequence 719 BP; 217 A; 171 C; 149 G; 182 T; 0 other;
 XX
 XX
 Query Match 56.0%; Score 14; DB 19; Length 719;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 2 caggtgcatgcgcca 15
 |||||||||||||
 DB 118 CAGGTGCATGCCTCA 105
 RESULT 19
 AAH48556
 ID AAH48556 standard; DNA: 943 BP.
 XX
 AC AAH48556;
 XX
 DT 20-SEP-2001 (first entry)
 XX
 DE Human fascin DNA fragment SEQ ID 6.
 XX
 KW Fascin: regulatory sequence; human; dendritic cell; antiviral; tumor;
 KW antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
 KW immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
 KW Creutzfeldt-Jakob disease; Alzheimer's disease; gene therapy;
 KW autoimmune disease; transplant rejection; ds.
 XX
 OS Homo sapiens.
 XX
 OS
 PN W0200151631-A2.
 XX
 PD 19-JUL-2001.
 XX
 PE 12-JAN-2001; 2001WO-EP00362.
 XX
 PF 13-JAN-2000; 2000DE-1001169.
 PR 02-MAR-2000; 2000DE-1010188.
 XX
 PA (RESK/) RESKE-KUNZ A.
 PA (ROSS/) ROSS X.
 PA (ROSS/) ROSS R.
 PA (BROS/) BROS M.
 XX
 PI Reske-Kunz A, Ross X, Ross R, Bros M;
 XX
 DR WPI: 2001-451858/48.
 XX
 PT New regulatory sequences from the fascin gene, useful for providing
 PT dendritic cell-specific expression of e.g. antigens, e.g. for
 PT vaccination against tumors and infections
 XX
 PS Claim 1c; Page 87; 117pp; German.
 XX
 CC This invention describes novel regulatory sequences (A) derived from
 CC human fascin that provide specific expression in dendritic cells (DC) and
 CC which have antiviral, antibacterial, antifungal, antiparasitic,
 CC anti-allergic, neurological, immunomodulatory and apoptotic activity.

(A) are used to regulate expression of antigens, immunoregulators, antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors and host cells that contain (A) are useful: (i) in vaccines against viruses, bacteria, fungi, parasites, tumors, allergens and plagues in Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of tumors, allergies, infections, autoimmune diseases and transplant rejection. They can also be provide specific expression of antigens and immunoregulators in DC; for isolation and identification of cell factors and cis-elements from regulatory sequences that mediate DC-specific expression; to determine the degree of maturity of DC and to block transcription factors, by providing binding sites in DC. (A) provide DC-specific expression of nucleic acid under their control, allowing a more specific regulation of the immune response and eliminating the long and laborious purification of DC (since a complete leucocyte population may be transformed), including transformation in vitro. This sequence represents a DNA fragment of human fascin which is described in the invention.

Sequence 943 BP; 135 A; 307 C; 340 G; 161 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 943;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gcatgcgcgcgcatt 20
DB 111 gcatgcgcgcgcatt 124
|||||

RESULT 20
AAC98831
ID AAC98831 standard; cDNA; 1190 BP.

AC AAC98831;
DT 09-MAR-2001 (first entry)

DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:59.

XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytostatic; neuroprotective;
KW neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antiinflammatory; cardiac; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative; ss.

XX Homo sapiens.

XX WO200055320-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000MO-US05989.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2000-579444/54.

XX P-PDB: AAB54066.

XX New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -

PS Claim 1; Page 537; 1379pp; English.

CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to

CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, neoplastic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiac and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.

Sequence 1190 BP; 269 A; 323 C; 329 G; 267 T; 2 other;

Query Match 56.0%; Score 14; DB 21; Length 1190;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ggtgcatgcgcgc 17
DB 1141 ggtgcatgcgcgc 1154
|||||

RESULT 21

AAB48555
ID AAB48555 standard; DNA; 1240 BP.

AC AAB48555;

DT 20-SEP-2001 (first entry)

DE Human fascin DNA fragment SEQ ID 5.

XX Fascin; regulatory sequence; human; dendritic cell; antiviral; tumor;
KW antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
KW immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
KW Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
KW autoimmune disease; transplant rejection; ds.

XX Homo sapiens.

XX WO200151631-A2.

XX 19-JUL-2001.

XX 12-JAN-2001; 2001WO-EP00362.

XX 13-JAN-2000; 2000DE-1001169.

XX 02-MAR-2000; 2000DE-1010186.

XX (RESK/) RESKE-KUNZ A.

XX (ROSS/) ROSS X.

XX (ROSS/) ROSS R.

XX (BROS/) BROS M.

XX Reske-Kunz A, Ross X, Ross R, Bros M;

XX WPI; 2001-451858/48.
XX New regulatory sequences from the fascin gene, useful for providing
PT dendritic cell-specific expression of e.g. antigens, e.g. for
PT vaccination against tumors and infections -
XX

Claim 1c; Page 87; 117pp; German

This invention describes novel regulatory sequences (A) derived from human fascin that provide specific expression in dendritic cells (DC) and which have antiviral, antibacterial, antifungal, antiparasitic, anti-allergic, neurological, immunomodulatory and apoptotic activity. (A) are used to regulate expression of antigens, immunoregulators, antisense sequences etc. in DC-specific fashion. Recombinant DNA, vector and host cells that contain (A) are useful: (i) in vaccines against viruses, bacteria, fungi, parasites, tumors, allergens and plaques in Creutzfeldt-Jakob and Alzheimer's disease; and (ii) for gene therapy of tumors, allergies, infections, autoimmune diseases and transplant rejection. They can also be used to provide specific expression of antigens and immunoregulators in DC; for isolation and identification of cell factors and cis-elements from regulatory sequences that mediate DC-specific expression; to determine the degree of maturity of DC and to block transcription factors, by providing binding sites in DC. (A) provide DC-specific expression of nucleic acid under their control, allowing a more specific regulation of the immune response and eliminating the long and laborious purification of DC (since a complete leucocyte population may be transformed), including transformation *in vitro*. This sequence represents a DNA fragment of human fascin which is described in the invention.

Query Match	56.0%;	Score 14;	DB 22;	Length 1240;
Best Local Similarity	100.0%;	Pred. No. 47;		
Matches 14;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

```
QY      7  gcatcgccagcat 20
         ||| ||| ||| ||| |||
Db     408  gcatcgccagcat 421
```

RESULT	22
AAZ56381/C	
ID	AAZ56381 standard; DNA; 1344 BP.

AC	AAZ56381;
XX	
DT	17-MAR-2000 (first entry)

Accession	Protein name	Length (aa)	Molecular weight (kDa)	pI	Isoelectric focusing
DE	Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:51	220	26.0	5.0	

kw Flagellin; flic; antigen; detection; ds

OS Escherichia coli

PN WO9961458-A1

PD 02-DEC-1999
xy

PE 21-MAY-1999; 99NO-A000383
XX

PR 21-MAY 1996, 2000 00000000
XX

XX

[illegible]

XX

PT bacterial strains in food, faeces, etc. -

Claim 3; Page 225; 245pp; English.

AAZ56331 to AAZ56399 represent nucleic acid molecules (1) encoding all or part of an *Escherichia coli* flagellin protein except a protein expressed by *E. coli* H1, H7, H12 or H48 type strains. The present invention also describes a method of detecting the presence of *E. coli*

CC of a particular H serotype in a sample, comprising specifically
CC hybridising a nucleic acid, preferably at least a pair, derived from a
CC flagellating gene, specific for a particular flagellin gene associated
CC with the H serotype, to any E.coli in the sample which contain the gene
CC and detecting any hybridised molecules, identifying the presence of that
CC serotype in the sample. (1) are useful for: (1) detecting the presence
CC of E. coli of H serotype in a sample by hybridising at least one or a
CC pair of (1) to any E. coli in the sample and detecting the hybridised
CC nucleic acid molecules; and (2) for detecting the presence of both O
CC and H-serotypes of E. coli by hybridising at least one or a pair of (1)
CC to any E. coli present in the sample and detecting the hybridised
CC nucleic acid molecules. (1) is particularly useful for detecting the
CC combination of O and H antigen. Hybridised (1) when using at least one
CC (1) is detected by southern blot analysis and, when using a pair of (1)
CC is detected by polymerase chain reaction (PCR). AAZ56399 to AAZ56440
CC represent primers used in the exemplification of the present invention.
XX
XX Sequence 1344 BP: 362 A; 337 C; 322 G; 323 T; 0 other:
SQ

Query Match	56.0%;	Score 14;	DB 21;	Length 1344;
Best Local Similarity	100.0%;	Pred. No. 47;		
Matches 14;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

```

QY      7  gcatcgccagcat 20
          |||||
Db      590 GCATGCCAGCAT 577

```

RESULT	23
AAH48554	
ID	AAH48554 standard; DNA; 1449 BP

AC	AAH48554;
XX	
DT	20-SEP-2001 (first entry)

DE Human fascin DNA fragment SEQ ID 4

KW Fascin, regulatory sequence; human; dendritic cell; antiviral; tumor;
KW antibacterial; antifungal; antiparasitic; anti-allergic; neurological
KW immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
KW Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
KW autoimmune disease; transplant rejection; ds.

OS Homo sapiens

PN W0200151631-A2

PD 19-JUL-2001

PF 12-JAN-2001; 2001WO-EP00362
XY

PR 13-JAN-2000; 2000DE-1001109
PR 02-MAR-2000; 2000DE-1010188

XX
PA (RESK/) RESKE-KUNZ A.

PA (RESK/) RESKE-KUNZ A
PA (ROSS/) ROSS X.
PA (ROSS/) ROSS R.
PA (BROS/) BROS M.

PI Reske-Kunz A, Ross X, Bros M

DR WPT; 2001-451858/48
XX

PN New regulatory sequences from the fascin gene, useful for providing
 PT dendritic cell-specific expression of e.g. antigens, e.g. for
 PT vaccination against tumors and infections

PS Claim 1c; Page 86-87; 117pp; German

AA This invention describes novel regulatory sequences (A) derived from
CC CC human fascin that provide specific expression in dendritic cells (DC) and
CC

CC which have antiviral, antibacterial, antifungal, antiparasitic,
CC anti-allergic, neurological, immunomodulatory and apoptotic activity.
CC (A) are used to regulate expression of antigens, immunoregulators,
CC antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors
CC and host cells that contain (A) are useful: (i) in vaccines against
CC viruses, bacteria, fungi, parasites, tumors, allergens and plaques in
CC Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of
CC tumors, allergies, infections, autoimmune diseases and transplant
CC rejection. They can also provide specific expression of antigens and
CC immunoregulators in DC; for isolation and identification of cell factors
CC and cis-elements from regulatory sequences that mediate DC-specific
CC transcription factors, by providing binding sites in DC. (A) provide
CC DC-specific expression of nucleic acid under their control, allowing a
CC more specific regulation of the immune response and eliminating the long
CC and laborious purification of DC (since a complete leucocyte population
CC may be transformed), including transformation in vitro. This sequence
CC represents a DNA fragment of human fascin which is described in the
CC invention.
CC
SQ Sequence 1449 BP: 232 A; 429 C; 532 G; 256 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1449;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 7 gcacgcacgacatt 20
Db 617 gcacgcacgacatt 630
|||||

RESULT 24
AAH48553
ID AAH48553 standard; DNA: 1619 BP.

AAH48553;

20-SEP-2001 (first entry)

Human fascin DNA fragment SEQ ID 3.

XX Fascin; regulatory sequence: human; dendritic cell; antiviral; tumor;
KW antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
KW immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
KW Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
KW autoimmune disease; transplant rejection; ds.
XX
OS Homo sapiens.
XX
PN WO200151631-A2.
XX
PD 19-JUL-2001.

12-JAN-2001; 2001WO-EP00362.

13-JAN-2000; 2000DE-1001169.

02-MAR-2000; 2000DE-1010188.

(RESK/) RESKE-KUNZ A.

(ROSS/) ROSS X.

(ROSS/) ROSS R.

(BROS/) BROS M.

Reske-Kunz A, Ross X, Ross R, Bros M;

WPI; 2001-451858/48.

XX New regulatory sequences from the fascin gene, useful for providing
PT dendritic cell-specific expression of e.g. antigens, e.g. for
PT vaccination against tumors and infections
XX
PS Claim 1c: Page 86; 117pp; German.

XX This invention describes novel regulatory sequences (A) derived from
CC human fascin that provide specific expression in dendritic cells (DC) and
CC which have antiviral, antibacterial, antifungal, antiparasitic,
CC anti-allergic, neurological, immunomodulatory and apoptotic activity.
CC (A) are used to regulate expression of antigens, immunoregulators,
CC antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors
CC and host cells that contain (A) are useful: (i) in vaccines against
CC viruses, bacteria, fungi, parasites, tumors, allergens and plaques in
CC Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of
CC tumors, allergies, infections, autoimmune diseases and transplant
CC rejection. They can also provide specific expression of antigens and
CC immunoregulators in DC; for isolation and identification of cell factors
CC and cis-elements from regulatory sequences that mediate DC-specific
CC expression; to determine the degree of maturity of DC and to block
CC transcription factors, by providing binding sites in DC. (A) provide
CC DC-specific expression of nucleic acid under their control, allowing a
CC more specific regulation of the immune response and eliminating the long
CC and laborious purification of DC (since a complete leucocyte population
CC may be transformed), including transformation in vitro. This sequence
CC represents a DNA fragment of human fascin which is described in the
CC invention.
CC
SQ Sequence 1619 BP: 278 A; 466 C; 583 G; 292 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1619;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 7 gcacgcacgacatt 20
Db 787 gcacgcacgacatt 800
|||||

RESULT 25
AAH48552
ID AAH48552 standard; DNA: 1934 BP.

AAH48552;

20-SEP-2001 (first entry)

Human fascin DNA fragment SEQ ID 2.

XX Fascin; regulatory sequence: human; dendritic cell; antiviral; tumor;
KW antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
KW immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
KW Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
KW autoimmune disease; transplant rejection; ds.
XX
OS Homo sapiens.
XX
PN WO200151631-A2.
XX
PD 19-JUL-2001.

12-JAN-2001; 2001WO-EP00362.

13-JAN-2000; 2000DE-1001169.

02-MAR-2000; 2000DE-1010188.

(RESK/) RESKE-KUNZ A.

(ROSS/) ROSS X.

(ROSS/) ROSS R.

(BROS/) BROS M.

Reske-Kunz A, Ross X, Ross R, Bros M;

WPI; 2001-451858/48.

XX New regulatory sequences from the fascin gene, useful for providing
PT dendritic cell-specific expression of e.g. antigens, e.g. for
PT vaccination against tumors and infections
XX
PS Claim 1c: Page 86; 117pp; German.

PT vaccination against tumors and infections -

XX PS Claim 1c: Page 85; 117pp; German.

XX CC This invention describes novel regulatory sequences (A) derived from

CC human fascin that provide specific expression in dendritic cells (DC) and

CC anti-allergic, neurological, immunomodulatory and apoptotic activity.

CC (A) are used to regulate expression of antigens, immunoregulators,

CC antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors

CC and host cells that contain (A) are useful: (i) in vaccines against

CC Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of

CC tumors, allergies, infections, autoimmune diseases and transplant

CC rejection. They can also be provide specific expression of antigens and

CC immunoregulators in DC; for isolation and identification of cell factors

CC and cis-elements from regulatory sequences that mediate DC-specific

CC expression; to determine the degree of maturity of DC and to block

CC transcription factors, by providing binding sites in DC. (A) provide

CC DC-specific expression of nucleic acid under their control, allowing a

CC more specific regulation of the immune response and eliminating the long

CC and laborious purification of DC (since a complete leucocyte population

CC may be transformed), including transformation in vitro. This sequence

CC represents a DNA fragment of human fascin which is described in the

CC invention.

CC XX

SQ Sequence 1934 BP; 344 A; 559 C; 647 G; 384 T; 0 other;

QY 7 gcatcgccagcact 20

DB 1102 gcatcgccagcact 1115

RESULT 26

AAA69762/c

ID AAA69762 standard; cDNA; 2017 BP.

XX AC AAA69762;

XX DT 07-NOV-2000 (first entry)

XX DE Human ovarian carcinoma antigen polynucleotide SEQ ID NO:72.

XX DE Human: ovarian carcinoma; ovarian cancer; therapy: diagnosis;

KW tumour antigen; Identification: cytostatic; gene therapy; vaccine; ss.

XX OS Homo sapiens.

XX PN WO200036107-A2.

XX PD 22-JUN-2000.

XX PF 17-DEC-1999; 99WO-US30270.

XX PR 17-DEC-1998; 98US-0215681.

PR 17-DEC-1998; 98US-0216003.

PR 23-JUN-1999; 99US-0338933.

PR 24-SEP-1999; 99US-0404879.

XX PA (CORI-) CORIXA CORP.

XX PI Mitcham JL, King GE, Algate PA, Frudakis TN;

XX DR WPI: 2000-431589/37.

XX PT Immunogenic portion of an ovarian carcinoma protein and the nucleic

PT acid encoding it, useful for the diagnosis, prevention and treatment of

PT cancer, preferably ovarian cancer -

XX PS Claim 1; Fig 2; 299pp; English.

XX CC The present invention describes an isolated polypeptide comprising an

CC immunogenic portion of an ovarian carcinoma protein (or its variants).

CC Ovarian carcinoma proteins, and polynucleotides encoding them, have

CC cytostatic activity and can be used in gene therapy and vaccines.

CC Ovarian carcinoma polypeptides, nucleic acids, antibodies and vaccines

CC are useful for the prevention, diagnosis and treatment of cancer.

CC preferably ovarian cancer. AAA69691 to AAA70077 and AAB12552 to AAB12557

CC represent human ovarian carcinoma polynucleotides and proteins used in

CC the exemplification of the present invention.

CC XX

SQ Sequence 2017 BP; 673 A; 440 C; 462 G; 442 T; 0 other;

QY 3 aggtgcacgcag 16

DB 964 AGCTGCATCGCAG 951

RESULT 27

AAT30684/c

ID AAT30684 standard; DNA; 2193 BP.

XX AC AAT30684;

XX DT 11-SEP-1996 (first entry)

XX DE Kaposi's sarcoma associated herpesvirus ORF22.

XX DE Kaposi's sarcoma associated herpesvirus; KSHV; gamma-2 herpesvirus;

KW therapy: diagnosis; antisense; triplex; vaccine; AIDS; ss.

XX OS Kaposi's sarcoma associated herpesvirus.

XX PN WO9615779-A1.

XX PD 30-MAY-1996.

XX PF 21-NOV-1995; 95WO-US15138.

XX PR 11-APR-1995; 95US-0420235.

PR 21-NOV-1994; 94US-0343101.

XX PA (UYCO) UNIV COLUMBIA NEW YORK.

XX PI Chang Y, Moore PS;

XX DR WPI: 1996-268320/27.

DR P-PSDB; AAR97832.

XX PT Herpes virus associated with Kaposi's sarcoma - also definitive DNA

PT sequences, useful for diagnosis of and to develop prods. for

PT treatment of Kaposi's sarcoma

XX PS Claim 17; Page 198-201; 277pp; English.

XX CC Lambda clone KS5 (AAT30681) has 17 protein-coding regions (AAT30682-98),

CC 15 of which are complete ORFs, including ORF22 (AAT30684) (nts 2093-

CC 4285), which codes for glycoprotein-H (AAR97832). KS5 is a fragment

CC of a newly identified human gamma-2 herpesvirus associated with

CC Kaposi's sarcoma (KS). The ORFs were identified by sequence and

CC positional homology to known herpesvirus sequences. Probes that bind

CC specifically to the KS associated herpesvirus can be used for KS

CC diagnosis. Antisense or triplex oligonucleotides are useful for

CC prophylaxis or treatment of KS, and the protein products (see also

CC AAR97830-46) of the 17 open reading frames are useful as vaccines.

SQ Sequence 2193 BP; 510 A; 597 C; 536 G; 550 T; 0 other;

Query Match 56.0%; Score 14; DB 17; Length 2193;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacacat 19
|||||
DB 46 TGCATGCCACGACAT 33

RESULT 28
AA16814/C
ID AA16814 standard; DNA; 2193 BP.
XX

AC AA16814;

DT 13-AUG-1996 (first entry)

DE Kaposi's sarcoma associated herpesvirus ORF22.

KW Kaposi's sarcoma; KSHV; lymphoma; AIDS; vaccine; diagnosis; therapy;
KM glycoprotein H; ss.

OS Kaposi's sarcoma associated herpesvirus.

PN WO9606159-A1.

PD 29-FEB-1996.

PF 11-AUG-1995; 95WO-US10194.

PR 11-APR-1995; 95US-0420325.

PR 18-AUG-1994; 94US-0292365.

PR 21-NOV-1994; 94US-0343100.

PA (UYCO) UNIV COLUMBIA NEW YORK.

PA (GRAN/) GRANT D E.

PA (VIEL/) VIELE L.

PI Chang Y, Moore PS, Grant DE, Vile L;

DR WPI; 1996-151362/15.

DR P-PSDB; AAR93608.

XX Herpesvirus DNA associated with Kaposi's sarcoma - also associated

XX vectors and proteins, used in detection and vaccination.

PS Claim 17; Page 211-214; 305pp; English.

XX Kaposi's sarcoma associated herpes virus (KHSV) clone KS5 (AA16806),

XX obtd. from a KS lesion genomic library, includes 15 complete ORFs and

XX 2 incomplete ORFs (AA16807-23) named according to their herpesvirus

XX salivari positional homologues. ORF22 (AA16814) codes for

XX glycoprotein H (AAR93608). KSHV DNA may be incorporated into a vector

XX and expressed in host cells to produce peptides useful in vaccines or

XX for raising antibodies. The DNA may itself be used to immunise a

XX subject, or used to design therapeutic antisense and triplex

XX molecules or diagnostic probes, or to raise transgenic animals.

SO Sequence 2193 BP; 510 A; 597 C; 536 G; 550 T; 0 other;

Query Match

Best Local Similarity 56.0%; Score 14; DB 17; Length 2193;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacacat 19
|||||
DB 46 TGCATGCCACGACAT 33

RESULT 29
AA20914/C
ID AA20914 standard; DNA; 2193 BP.
XX

AC AA20914;

DT 02-DEC-1999 (first entry)

DE Nucleotide sequence of gH glycoprotein of HHV-8.

KW vaccine; antibody; Kaposi's sarcoma; glycoprotein; detection;

KW HHV-8; Human Herpesvirus type 8; ds.

OS Human Herpes Virus type 8.

PN WO945948-A1.

PD 16-SEP-1999.

PF 12-MAR-1999; 99WO-US05464.

PR 12-MAR-1998; 98US-0077670.

PA (UYPE-) UNIV PENNSYLVANIA.

PI Whitbeck JC, Eisenberg RJ, Cohen GH;

DR WPI; 1999-551211/46.

DR P-PSDB; AAY42395.

XX New nucleic acid molecules encoding human herpesvirus type glycoprotein

XX H and glycoprotein L polypeptides for diagnosis and treatment of

XX Kaposi's Sarcoma

XX Claim 3; Fig 1; 75pp; English.

XX This is the nucleotide sequence of the human herpes virus type 8

XX glycoprotein H. This sequence and it's encoded protein (AAY42395) can be

XX used in the development of vaccines, antibodies and other preparations.

XX The vaccine is useful as a therapeutic agent for prevention of HHV-8

XX disease and as a therapeutic agent for treatment of HHV-8 infection

XX in order to boost the immune response in the animal. Administration of

XX a pure preparation of soluble HHV-8 gH/gL complex or at least one

XX polynucleotide encoding the soluble complex is useful for preventing or

XX treating an HHV-8 infection in a human.

XX Diagnosing an HHV-8 infection in a human is carried out by contacting

XX a biological sample with an antibody which specifically binds to either

XX glycoprotein or in complex form. If binding of the antibody occurs then

XX the human has HHV-8 infection.

XX Alternatively the infection is diagnosed by addition of one of the

XX HHV-8 glycoproteins or complex and determining whether specific binding

XX of the protein to an antibody in the biological sample occurs. A further

XX diagnostic method involves contacting the biological sample with an

XX HHV-8 glycoprotein H polynucleotide or HHV-8 glycoprotein L

XX sample occurs.

SO Sequence 2193 BP; 510 A; 597 C; 536 G; 550 T; 0 other;

Query Match

Best Local Similarity 56.0%; Score 14; DB 20; Length 2193;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacacat 19
|||||
DB 46 TGCATGCCACGACAT 33

RESULT 30

AAH48551

ID AAH48551 standard; DNA; 3069 BP.

AAH48551;
 20-SEP-2001 (first entry)
 Human fascin DNA fragment SEQ ID 1.
 Fascin: regulatory sequence: human; dendritic cell; antiviral; tumor;
 antibacterial; antifungal; antiparasitic; anti-allergic; neurotoxic;
 immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
 Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
 autoimmune disease; transplant rejection; ds.
 Homo sapiens.
 WO200151631-A2.
 19-JUL-2001.
 12-JAN-2001: 2001WO-EP00362.
 13-JAN-2000: 2000DE-1001169.
 02-MAR-2000: 2000DE-1010188.
 (RESK/) RESKE-KUNZ A.
 (ROSS/) ROSS X.
 (ROSS/) ROSS R.
 (BROS/) BROS M.
 Reske-Kunz A, Ross X, Ross R, Bros M;
 WPI: 2001-451858/48.
 New regulatory sequences from the fascin gene, useful for providing
 dendritic cell-specific expression of e.g. antigens, e.g. for
 vaccination against tumors and infections
 Claim 1a: Page 84-85; 117pp; German.
 This invention describes novel regulatory sequences (A) derived from
 human fascin that provide specific expression in dendritic cells (DC) and
 which have antiviral, antibacterial, antifungal, antiparasitic,
 anti-allergic, neurological, immunomodulatory and apoptotic activity.
 (A) are used to regulate expression of antigens, immunoregulators,
 antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors
 and host cells that contain (A) are useful: (i) in vaccines against
 viruses, bacteria, fungi, parasites, tumors, allergens and plaques in
 Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of
 tumors, allergies, infections, autoimmune diseases and transplant
 rejection. They can also be provide specific expression of cell factors
 and immunoregulators in DC; for isolation and identification of cell factors
 and cis-elements from regulatory sequences that mediate DC-specific
 expression; to determine the degree of maturity of DC and to block
 transcription factors, by providing binding sites in DC. (A) provide a
 DC-specific expression of nucleic acid under their control, allowing a
 more specific regulation of the immune response and eliminating the long
 and laborious purification of DC (since a complete leucocyte population
 may be transformed), including transformation in vitro. This sequence
 represents a DNA fragment of human fascin which is described in the
 invention.
 Sequence 3069 BP: 607 A; 867 C; 993 G; 602 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 3069;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gcatcgccagcatt 20
 |||||
 DB 2237 gcatcgccagcatt 2250

RESULT 31

AAZ39011/C
 ID AAZ39011 standard; cDNA; 3593 BP.
 AAZ39011;
 28-FEB-2000 (first entry)
 Mouse Eseq2 coding sequence.
 Mouse; murine; Eseq2; endocytosis; vesicular trafficking;
 regulation; actin cytoskeleton; detection; cancer; infection;
 EH-domain and SH3-domain regulator of endocytosis; anticancer;
 antiproliferative; antiviral; ss.
 Mus sp.
 WO9955728-A2.
 04-NOV-1999.
 27-APR-1999: 99WO-CA00375.
 27-APR-1998: 98CA-2230201.
 05-FEB-1999: 99US-0118739.
 (HSCR-) HSC RES & DEV LP.
 Egan SE, Wang W, Sengar A;
 WPI: 2000-052802/04.
 P-PSDB: AAY57445.
 New nucleic acid encoding Eseq1 and 2 proteins, involved in regulation
 of endocytosis, used e.g. for treating cancer or preventing viral
 infection
 Claim 25: Page 46-48; 99pp; English.
 The present sequence encodes mouse Eseq2. The present invention
 specifically describes mammalian Eseq1 and 2 proteins (I) and their splice
 variants (Eseq = EH-domain and SH3-domain regulator of endocytosis). (I)
 are involved in regulation of clathrin-mediated endocytosis (as a complex
 with Eps15 protein), vesicular trafficking and actin cytoskeleton.
 Generally (I) (or its (ant)agonists, mimetics, fragments and inactive
 mutants): (I)-specific antibodies (Ab): sequences antisense to the (I)
 polynucleotide; agents that downregulate expression of Eseq genes or
 antagonists of an Eseq binding partner are used to treat diseases
 associated with undesirable endocytosis and resulting changes in cellular
 function. Particularly overexpression of Eseq1 is used to block
 clathrin-mediated endocytosis in vivo or in cell cultures, while
 administration of (I) is used to promote endocytosis of selected cells.
 CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
 cells that can be stimulated to proliferate by a growth factor receptor;
 and similar compounds (also inactive Eseq mutants) can be used to prevent
 viral infection. Endocytosis may also be regulated, in vivo or in cell
 cultures, by forming an Eseq-Eps15 complex, then binding dynamin to the
 complex. Generally conditions that can be treated include cancer;
 CC abnormal cell division or migration; viral infection; or abnormal
 receptor signalling, tissue development or synaptic transmission.
 Sequence 3593 BP: 1171 A; 783 C; 850 G; 789 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 3593;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 aggtgcatgcccag 16
 |||||
 DB 923 AGGTGATGCGCCAG 910

RESULT 32

AAH48581
ID AAH48581 standard; DNA: 4100 BP.
XX
AC AAH48581;
XX
DT 20-SEP-2001 (first entry)
XX
DE Human fascin DNA fragment SEQ ID 33.
XX
KW Fascin; regulatory sequence; human; dendritic cell; antiviral; tumor;
KM antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
KW immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
KM Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
KW autoimmune disease; transplant rejection; ds.
XX
OS Homo sapiens.
XX
PN WO200151631-A2.
XX
PD 19-JUL-2001.
XX
PF 12-JAN-2001: 2001WO-EP00362.
XX
PR 13-JAN-2000: 2000DE-1001169.
PR 02-MAR-2000: 2000DE-1010188.
XX
PA (RESK/) RESKE-KUNZ A.
PA (ROSS/) ROSS X.
PA (ROSS/) ROSS R.
PA (BROS/) BROS M.
XX
PI Reske-Kunz A, Ross X, Ross R, Bros M;
XX
DR WPI: 2001-451858/48.
XX
PT New regulatory sequences from the fascin gene, useful for providing
PT dendritic cell-specific expression of e.g. antigens, e.g. for
PT vaccination against tumors and infections
XX
PS Example 1; Page 102-103; 117pp; German.
XX
CC This invention describes novel regulatory sequences (A) derived from
CC human fascin that provide specific expression in dendritic cells (DC) and
CC which have antiviral, antibacterial, antifungal, antiparasitic,
CC anti-allergic, neurological, immunomodulatory and apoptotic activity.
CC (A) are used to regulate expression of antigens, immunoregulators,
CC antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors
CC and host cells that contain (A) are useful: (i) in vaccines against
CC Creutzfeld-Jakob and Alzheimer's disease; tumors, allergens and plaques in
CC tumors, allergies, infections, autoimmune diseases and transplant
CC rejection. They can also be provide specific expression of antigens and
CC immunoregulators in DC; for isolation and identification of cell factors
CC and cis-elements from regulatory sequences that mediate DC-specific
CC expression; to determine the degree of maturity of DC and to block
CC DC-transcription factors, by providing binding sites in DC. (A) provide
CC DC-specific expression of nucleic acid under their control, allowing a
CC more specific purification of the immune response and eliminating the long
CC and laborious purification of DC (since a complete leucocyte population
CC may be transformed), including transformation in vitro. This sequence
CC represents a DNA fragment of human fascin which is described in the
CC invention.
XX
XX Sequence 4100 BP; 769 A; 1251 C; 1339 G; 741 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 4100;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gcacgcgcagcatt 20
DB 2237 gcacgcgcagcatt 2250

RESUL T 33
AAZ49334
ID AAZ49334 standard; cDNA: 4189 BP.
XX
AC AAZ49334;
XX
DT 14-MAR-2000 (first entry)
XX
DE Murine multidrug resistance-1 (MDR-1) cDNA.
XX
KW Multidrug resistance; MDR-1; P-glycoprotein;
KM transmembrane efflux pump; haematopoietic stem cell; transduction;
KW bone marrow transplantation; chemotherapy; radiation therapy; cancer;
KW gene therapy; gene replacement; genetic defect; thalassaemia;
KW Gaucher's disease; sickle cell anaemia; leukaemia; ex vivo expansion;
KW cytokine; ds.
XX
OS Mus sp.
XX
XX Key Location/Qualifiers
XX FT 1..3831
XX CDS /*tag= a
XX FT /*product= "Murine MDR-1 protein"
XX
XX WO9961589-A2.
XX
XX 02-DEC-1999.
XX
XX 27-MAY-1999: 99WO-US11825.
XX
XX 28-MAY-1998: 98US-0086988.
XX
XX (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX
XX Sorrentino B, Bunting K;
XX
XX WPI: 2000-072615/06.
XX
XX P-PSDB; AAY58188.
XX
XX Ex vivo expansion of hematopoietic stem cells transduced with a
XX PT sequence encoding human multidrug resistance-1, used for bone marrow
XX PT transplantation
XX
XX Disclosure: Page 90-93; 113pp; English.
XX
CC This sequence represents cDNA encoding murine multidrug
CC resistance protein MDR-1. MDR-1 is a transmembrane
CC efflux pump, responsible for the export of drugs from cells,
CC particularly cancer cells. The invention relates to transducing
CC haematopoietic stem cells with nucleic acid encoding an MDR protein
CC and culturing the modified cells. The modified haematopoietic stem
CC cells are useful in bone marrow transplantation (to reconstitute
CC haematopoietic systems in patients who have undergone chemotherapy or
CC radiation therapy) and in ex vivo gene therapy of genetic defects in
CC cells derived from haematopoietic stem cells, e.g., thalassaemia,
CC Gaucher's disease, sickle cell anaemia or leukaemia. The modified
CC cells can also be used to identify factors involved in regulating
CC proliferation and differentiation in haematopoietic stem cells.
CC Haematopoietic stem cells that express MDR-1 will be protected against
CC chemotherapeutic agents, so can be engrafted while the patient is
CC undergoing chemotherapy. Expansion of (rare) haematopoietic stem cells
CC provides sufficient cells to permit standard biochemical analysis.
CC Overexpression of MDR-1 allows cytokine-driven expansion of
CC haematopoietic stem cells by at least 10-fold compared with a maximum
CC of 4-fold in known procedures.
XX
XX Sequence 4189 BP; 1204 A; 875 C; 1028 G; 1082 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 4189;
Best Local Similarity 100.0%; Pred. No. 46;

Matches 14: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
 OY 12 gccagcattcgat 25
 |||||
 DB 880 gccagcattcgat 893

RESULT 34

AAQ38950
 ID AAQ38950 standard; DNA: 4313 BP.

XX
 AC AAQ38950:

XX 28-JUL-1993 (first entry)

XX Mouse multidrug resistance sequence.

XX mdr gene; Lambda DR11 clone; ss.

XX Mus musculus.

XX Key Location/Qualifiers
 FH 110..3940
 FT CDS //tag= a
 FT //phenotype= multidrug_resistance

XX US198344-A.

XX 30-MAR-1993.

XX 15-JUL-1986; 86US-0885951.

XX 15-JUL-1986; 86US-0885951.

XX 06-FEB-1991; 91US-0652311.

XX (MAST) MASSACHUSETTS INST TECHNOLOGY.

XX Croop JM, Gros P, Housman DE;

XX WPI: 1993-126077/15.

XX P-PSDB; AAR35199.

XX DNA sequence which confers multi-drug resistance on sensitive

XX mammalian cells - used to preserve bone marrow cells during

XX chemotherapy to prevent infection

XX Claim 1: Fig 8: 22pp; English.

XX A cDNA library was constructed from mRNA which had been isolated

XX from a drug-sensitive mouse cell-line. Two mouse cDNA molecules

XX complementary to the mRNA species encoded by 2 related but distinct

XX mdr genes were isolated and cloned. One of the cDNA clones (lambda

XX DR11) is a full-length cDNA clone for one member of the mdr gene

XX family. When incorporated into prokaryotic expression vector pORFEX4

XX (which allows high levels of transcription of the cDNA when

XX introduced into mammalian cells), the clone was shown to confer the

XX multidrug resistance phenotype upon transfection into drug-sensitive

XX mammalian cells.

XX Sequence 4313 BP; 1241 A; 904 C; 1058 G; 1110 T; 0 other;

XX Query Match 56.0%; Score 14; DB 14; Length 4313;

XX Best Local Similarity 100.0%; Pred. No. 46;

XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ID AA239010 standard; cDNA: 4625 BP.

XX AA239010;

XX 28-FEB-2000 (first entry)

XX Mouse Ese2 full length cDNA sequence.

XX Mouse; murine; Ese1: Ese2; endocytosis; vesicular trafficking;

XX regulation; actin cytoskeleton; detection; cancer; infection;

XX EH-domain and SH3-domain regulator of endocytosis; anticancer;

XX antiproliferative; antiviral; ss.

XX Mus sp.

XX W09955728-A2.

XX 04-NOV-1999.

XX 27-APR-1999; 99WO-CA00375.

XX 27-APR-1998; 98CA-2230201.

XX 05-FEB-1999; 99US-0118739.

XX (HSCR-) HSC RES & DEV LP.

XX Egan SE, Wang W, Sengar A;

XX WPI: 2000-052802/04.

XX P-PSDB: AAY57445.

XX New nucleic acid encoding Ese1 and 2 proteins, involved in regulation

XX of endocytosis, used e.g. for treating cancer or preventing viral

XX infection -

XX Claim 25: Page 43-46; 99pp; English.

XX The present sequence encodes mouse Ese2. The present invention

XX specifically describes mammalian Ese1 and 2 proteins (I) and their splice

XX variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I)

XX are involved in regulation of clathrin-mediated endocytosis (as a complex

XX with Eps15 protein), vesicular trafficking and actin cytoskeleton.

XX Generally (I) (or its (antagonists (Ab); sequences antisense to the (I)

XX mutants); (I)-specific antibodies (Ab); sequences antisense to the (I)

XX polynucleotide; agents that downregulate expression of Ese genes or

XX antagonists of an Ese binding partner are used to treat diseases or

XX associated with undesirable endocytosis and resulting changes in cellular

XX function. Particularly overexpression of Ese1 is used to block

XX clathrin-mediated endocytosis in vivo or in cell cultures, while

XX administration of (I) is used to promote endocytosis of selected cells.

XX (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of

XX cells that can be stimulated to proliferate by a growth factor receptor;

XX and similar compounds (also inactive Ese mutants) can be used to prevent

XX viral infection. Endocytosis may also be regulated, in vivo or in cell

XX cultures, by forming an Ese-Eps15 complex, then binding dynamin to the

XX complex. Generally conditions that can be treated include cancer;

XX abnormal cell division or migration; viral infection; or abnormal

XX receptor signalling, tissue development or synaptic transmission.

XX Sequence 4625 BP; 1453 A; 1009 C; 1104 G; 1059 T; 0 other;

XX Query Match 56.0%; Score 14; DB 21; Length 4625;

XX Best Local Similarity 100.0%; Pred. No. 45;

XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 agtgcacgcgcag 16

DB 1173 AGTGCATCGCCAG 1160

RESULT 36

AA239027/c

ID AA239027 standard; cDNA; 4975 BP.
 AA239027;
 28-FEB-2000 (first entry)
 Mouse Ese2L coding sequence.
 Mouse: murine; Ese1; Ese2; endocytosis; vesicular trafficking;
 regulation; actin cytoskeleton; detection; cancer; infection;
 EH-domain and SH3-domain regulator of endocytosis; anticancer;
 antiproliferative; antiviral; ss.
 Mus sp.
 W09955728-A2.
 04-NOV-1999.
 27-APR-1999; 99WO-CA00375.
 27-APR-1998; 98CA-2230201.
 05-FEB-1999; 99NS-0118739.
 (HSCR-) HSC RES & DEV LP.
 Egan SE, Wang W, Sengar A;
 WPI: 2000-052802/04.
 P-PSDB: AAV57450.
 New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
 of endocytosis, used e.g. for treating cancer or preventing viral
 infection -
 Claim 25; Page 66-69; 99pp; English.
 The present invention specifically describes mammalian Ese1 and 2
 proteins (I) and their splice variants (Ese = EH-domain and SH3-domain
 regulator of endocytosis). (I) are involved in regulation of clathrin-
 mediated endocytosis (as a complex with Esp15 protein), vesicular
 trafficking and actin cytoskeleton. Generally (I) (or its antagonists;
 mimetics, fragments and inactive mutants); (II)-specific antibodies (Ab);
 sequences antisense to the (I) polynucleotide; agents that downregulate
 expression of Ese genes or antagonists of an Ese binding partner are
 used to treat diseases associated with undesirable endocytosis and
 resulting changes in cellular function. Particularly overexpression of
 Ese1 is used to block clathrin-mediated endocytosis in vivo or in cell
 cultures, while administration of (I) is used to promote endocytosis of
 selected cells. (Antagonists of (I) or Ab are used to suppress abnormal
 proliferation of cells that can be stimulated to proliferate by a growth
 factor receptor; and similar compounds (also inactive Ese mutants) can be
 used to prevent viral infection. Endocytosis may also be regulated, in
 vivo or in cell cultures, by forming an Ese-Esp15 complex, then binding
 dynamo to the complex. Generally conditions that can be treated include
 cancer; abnormal cell division or migration; viral infection; or abnormal
 receptor signalling, tissue development or synaptic transmission. The
 present sequence represents mouse Ese2L coding sequence.
 Sequence 4975 BP; 1540 A; 1166 C; 1201 G; 1068 T. A other.

```

Query Match: 56.0%; Score 14; DB 21; length 4975;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 aggtlcatcgccag 16
      ||| ||||| |||||
Db 923 AGCTGCATCGCCAG 910

```

ID	AAV35235	standard; DNA; 5319 BP.
XX		
AC	AAV35235;	
XX		
DJ	28-SEP-1998	(first entry)
XX		
DE	R. prowazekii	S-layer protein genomic DNA.
KW	Surface layer protein; S-layer; vaccine; typhus; spotted fever;	
KM	infection; diagnosis; disease; ds.	
XX		
OS	Rickettsia prowazekii.	
XX		
FH	Key	Location/Qualifiers
FT	-35_signal	340..345
FT		/*tag= a
FT	-10_signal	363..368
FT		/*tag= b
FT	RBS	379..386
FT		/*tag= c
FT	CDS	391..5229
FT		/*tag= d
FT	stem_loop	/product= S-layer protein
FT		5270..5306
XX		/*tag= e
PN	US5783441-A.	
XX		
PD	21-JUL-1998.	
XX		
PF	20-DEC-1993;	93US-0169927.
XX		
PR	20-DEC-1993;	93US-0169927.
PR	09-AUG-1991;	91US-0742128.
XX		
PA	(USNA)	US SEC OF NAVY.
XX		
PI	Carl M, Ching W, Dasch GA, Dobson ME;	
XX		
DR	WPI: 1998-427031/36.	
DR	P-PSDB; AAW65088.	
XX		
PT	Recombinant DNA encoding Rickettsia surface layer proteins - useful	
PT	for diagnosing typhus and spotted fever and for preparing vaccines	
PT	against them	
XX		
PS	Claim 1; Column 11-24; 20pp; English.	
XX		
CC	This sequence encodes the Surface layer (S-layer) protein from	
CC	R. prowazekii strain Breinl. This sequence is useful for vaccination	
CC	against typhus and spotted fever rickettsial infection or for diagnosing	
CC	diseases caused by these bacteria. The surface layer protein antigens can	
CC	be produced recombinantly in large quantities.	
XX		
XX		
3Q	Sequence 5319 BP; 1815 A; 766 C; 983 G; 1755 T; 0 other	

```

Query Match          56.0%; Score 14; DB 19; Length 5319;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 catcgccagcattt 21
        |||
Db       2221 CATCGCCAGCATT 2208

RESULT 38
AAV21186/c
ID      AAV21186 standard; DNA; 5676 BP.
XX
XX
AC      AAV21186;
XX
XX      24-JUL-1998 (first entry)
XX      XX

```

```

XX DE Amycolatopsis mediterranei strain wt3136 5.7 kb KpnI DNA fragment.
XX
XX KM Amycolatopsis mediterranei rifamycin synthesis; gene cluster:
XX KW polyketide synthase; actinomycete; ansamycin; ds.
XX OS Amycolatopsis mediterranei.
XX
XX FH Key Location/Qualifiers
XX FT CDS 3..5676
XX FT /*tag= a
XX FT /product= "polyketide synthase"
XX FT /note= "no stop codon given"
XX
XX PN MO9807868-A1.
XX
XX PD 26-FEB-1998.
XX
XX PF 18-AUG-1997; 97WO-EP04495.
XX
XX PR 20-AUG-1996; 96EP-0810551.
XX
XX PA (NOVS ) NOVARTIS AG.
XX
XX PI Engel N, Schupp T, Toupet C;
XX
XX DR WPI: 1998-169172/15.
XX DR P-PSDB: AAM52844.
XX
XX PT Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
XX PS to produce rifamycin and rifamycin analogues
XX
XX PS Claim 4: Page 38-43; 205pp; English.
XX
XX CC The present sequence represents Amycolatopsis mediterranei strain wt3136
XX CC 5.7 kb KpnI fragment DNA, from the present invention. The present
XX CC invention describes a Amycolatopsis mediterranei rifamycin synthesis
XX CC gene cluster DNA fragment comprising a DNA region involved directly or
XX CC indirectly in the gene cluster responsible for rifamycin synthesis,
XX CC including the adjacent DNA regions to the right and left which, by
XX CC reason of their function in connection with rifamycin biosynthesis,
XX CC qualify as constituents of this rifamycin gene cluster, and functional
XX CC fragments, derivatives or constituents of these. The Amycolatopsis
XX CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
XX CC for producing rifamycin, rifamycin analogues or precursors. It can also
XX CC be used for inactivating or modifying genes involved in ansamycin or
XX CC rifamycin biosynthesis. The DNA can be used for constructing mutant
XX CC actinomycetes strains from which the natural rifamycin or ansamycin
XX CC biosynthesis gene cluster has been partly or completely deleted. The
XX CC DNA fragment can be used for assembling a library of polyketide
XX CC synthases, which can be used for assembling a library of polyketides.
XX CC A hybridisation probe of the invention can be used for identifying DNA
XX CC fragments involved in the biosynthesis of ansamycins.
XX
XX SQ Sequence 5676 BP; 674 A; 1990 C; 2205 G; 807 T; 0 other:

```

```

Query Match 56.0%; Score 14; DB 19; Length 5676;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 2 caggtgcatcgcca 15
   |||
DB 3923 CAGGTGTCATCGCCA 3910

```

```

RESULT 39
AA239026/C
ID AA239026 standard; cDNA: 6014 BP.
XX
XX AC AA239026;
XX
XX DT 28-FEB-2000 (first entry)

```

```

XX DE Mouse Ese2L cDNA sequence.
XX
XX KM Mouse; murine; Ese1; Ese2; endocytosis; vesicular trafficking;
XX KW regulation; actin cytoskeleton; detection; cancer; infection;
XX KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
XX KW antiproliferative; antiviral; ss.
XX
XX OS Mus sp.
XX
XX PN W09955728-A2.
XX
XX PD 04-NOV-1999.
XX
XX PF 27-APR-1999; 99MO-CA00375.
XX
XX PR 27-APR-1998; 98CA-2230201.
XX PR 05-FEB-1999; 99US-0118739.
XX
XX PA (HSCR-) HSC RES & DEV LP.
XX
XX PI Egan SE, Wang W, Sengar A;
XX
XX DR WPI: 2000-052802/04.
XX DR P-PSDB: AAY57450.
XX
XX PT New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
XX PT of endocytosis, used e.g. for treating cancer or preventing viral
XX PT infection.
XX
XX PS Claim 25; Page 63-66; 99pp; English.
XX
XX CC The present invention specifically describes mammalian Ese1 and 2
XX CC proteins (I) and their splice variants (Ese = EH-domain and SH3-domain
XX CC regulator of endocytosis). (I) are involved in regulation of clathrin-
XX CC mediated endocytosis (as a complex with Eps15 protein), vesicular
XX CC trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,
XX CC mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);
XX CC sequences antisense to the (I) polynucleotide; agents that downregulate
XX CC expression of Ese genes or antagonists of an Ese binding partner are
XX CC used to treat diseases associated with undesirable endocytosis and
XX CC resulting changes in cellular function. Particularly overexpression of
XX CC Ese1 is used to block clathrin-mediated endocytosis in vivo or in cell
XX CC cultures, while administration of (I) is used to promote endocytosis of
XX CC selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal
XX CC proliferation of cells that can be stimulated to proliferate by a growth
XX CC factor receptor, and similar compounds (also inactive Ese mutants) can be
XX CC used to prevent viral infection. Endocytosis may also be regulated, in
XX CC vivo or in cell cultures, by forming an Ese-Eps15 complex, then binding
XX CC dynam to the complex. Generally conditions that can be treated include
XX CC cancer; abnormal cell division or migration; viral infection; or abnormal
XX CC receptor signalling, tissue development or synaptic transmission. The
XX CC present sequence represents mouse Ese2L cDNA sequence.
XX
XX SQ Sequence 6014 BP; 1833 A; 1389 C; 1454 G; 1338 T; 0 other:

```

```

Query Match 56.0%; Score 14; DB 21; Length 6014;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 3 aggtgcatcgccag 16
   |||
DB 1173 AGGTGTCATCGCCAG 1160

```

```

RESULT 40
AAH48621
ID AAH48621 standard; DNA: 13055 BP.
XX
XX AC AAH48621;
XX
XX DT 20-SEP-2001 (first entry)

```

```

XX DE Human fascin DNA fragment #1.
XX XX
XX Fascin: regulatory sequence; human; dendritic cell; antiviral; tumor;
XX antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
XX immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
XX Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
XX autoimmune disease; transplant rejection; ds.
XX OS Homo sapiens.
XX PN MO200151631-A2.
XX PD 19-JUL-2001.
XX PF 12-JAN-2001; 2001WO-EP00362.
XX PR 13-JAN-2000; 2000DE-1001169.
XX PR 02-MAR-2000; 2000DE-1010188.
XX PA (RESK/) RESKE-KUNZ A.
XX PA (ROSS/) ROSS X.
XX PA (ROSS/) ROSS R.
XX PA (BROS/) BROS M.
XX PI Reske-Kunz A, Ross X, Ross R, Bros M;
XX DR WPI: 2001-451858/48.
XX PT New regulatory sequences from the fascin gene, useful for providing
XX PT dendritic cell-specific expression of e.g. antigens, e.g. for
XX PT vaccination against tumors and infections
XX PS Example 1; Fig 2; 117pp; German.
XX XX
XX This invention describes novel regulatory sequences (A) derived from
XX CC human fascin that provide specific expression in dendritic cells (DC) and
XX CC which have antiviral, antibacterial, antifungal, antiparasitic,
XX CC anti-allergic, neurological, immunomodulatory and apoptotic activity.
XX CC (A) are used to regulate expression of antigens, immunoregulators,
XX CC antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors
XX CC and host cells that contain (A) are useful: (i) in vaccines against
XX CC viruses, bacteria, fungi, parasites, tumors, allergens and plaques in
XX CC Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of
XX CC tumors, allergies, infections, autoimmune diseases and transplant
XX CC rejection. They can also be provide specific expression of antigens and
XX CC immunoregulators in DC; for isolation and identification of cell factors
XX CC and cis-elements from regulatory sequences that mediate DC-specific
XX CC expression; to determine the degree of maturity of DC and to block
XX CC transcription factors, by providing binding sites in DC. (A) provide
XX CC DC-specific expression of nucleic acid under their control, allowing a
XX CC more specific regulation of the immune response and eliminating the long
XX CC and laborious purification of DC (since a complete leucocyte population
XX CC may be transformed), including transformation in vitro. This sequence
XX CC represents a DNA fragment of human fascin which is described in the
XX CC invention.
XX SO Sequence 13055 BP; 2372 A; 3903 C; 4097 G; 2683 T; 0 other;

```

```

Query Match 56.0%; Score 14; DB 22; Length 13055;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 7 gcacgcacgacatt 20
   |||
Db 2237 gcacgcacgacatt 2250

```

```

RESULT 41
AAH48620
ID AAH48620 standard; DNA: 16951 BP.
XX

```

```

AC AAH48620;
XX XX
XX 20-SEP-2001 (first entry)
XX XX
XX Human fascin DNA fragment SEQ ID 72.
XX DE
XX XX
XX Fascin: regulatory sequence; human; dendritic cell; antiviral; tumor;
XX antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
XX immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
XX Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
XX autoimmune disease; transplant rejection; ds.
XX OS Homo sapiens.
XX PN MO200151631-A2.
XX PD 19-JUL-2001.
XX PF 12-JAN-2001; 2001WO-EP00362.
XX PR 13-JAN-2000; 2000DE-1001169.
XX PR 02-MAR-2000; 2000DE-1010188.
XX PA (RESK/) RESKE-KUNZ A.
XX PA (ROSS/) ROSS X.
XX PA (ROSS/) ROSS R.
XX PA (BROS/) BROS M.
XX PI Reske-Kunz A, Ross X, Ross R, Bros M;
XX DR WPI: 2001-451858/48.
XX PT New regulatory sequences from the fascin gene, useful for providing
XX PT dendritic cell-specific expression of e.g. antigens, e.g. for
XX PT vaccination against tumors and infections
XX PS Claim 1a; Page 112-117; 117pp; German.
XX XX
XX This invention describes novel regulatory sequences (A) derived from
XX CC human fascin that provide specific expression in dendritic cells (DC) and
XX CC which have antiviral, antibacterial, antifungal, antiparasitic,
XX CC anti-allergic, neurological, immunomodulatory and apoptotic activity.
XX CC (A) are used to regulate expression of antigens, immunoregulators,
XX CC antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors
XX CC and host cells that contain (A) are useful: (i) in vaccines against
XX CC viruses, bacteria, fungi, parasites, tumors, allergens and plaques in
XX CC Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of
XX CC tumors, allergies, infections, autoimmune diseases and transplant
XX CC rejection. They can also be provide specific expression of antigens and
XX CC immunoregulators in DC; for isolation and identification of cell factors
XX CC and cis-elements from regulatory sequences that mediate DC-specific
XX CC expression; to determine the degree of maturity of DC and to block
XX CC transcription factors, by providing binding sites in DC. (A) provide
XX CC DC-specific expression of nucleic acid under their control, allowing a
XX CC more specific regulation of the immune response and eliminating the long
XX CC and laborious purification of DC (since a complete leucocyte population
XX CC may be transformed), including transformation in vitro. This sequence
XX CC represents a DNA fragment of human fascin which is described in the
XX CC invention.
XX SO Sequence 16951 BP; 3224 A; 4928 C; 5200 G; 3599 T; 0 other;

```

```

Query Match 56.0%; Score 14; DB 22; Length 16951;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 7 gcacgcacgacatt 20
   |||
Db 2237 gcacgcacgacatt 2250

```

```

RESULT 42

```


XX WPI: 1999-069741/06.

DR Kaposi's sarcoma-associated herpes virus nucleic acid - encodes
XX dihydrofolate reductase and is useful for treatment, prophylaxis
PT or diagnosis of Kaposi's sarcoma
XX

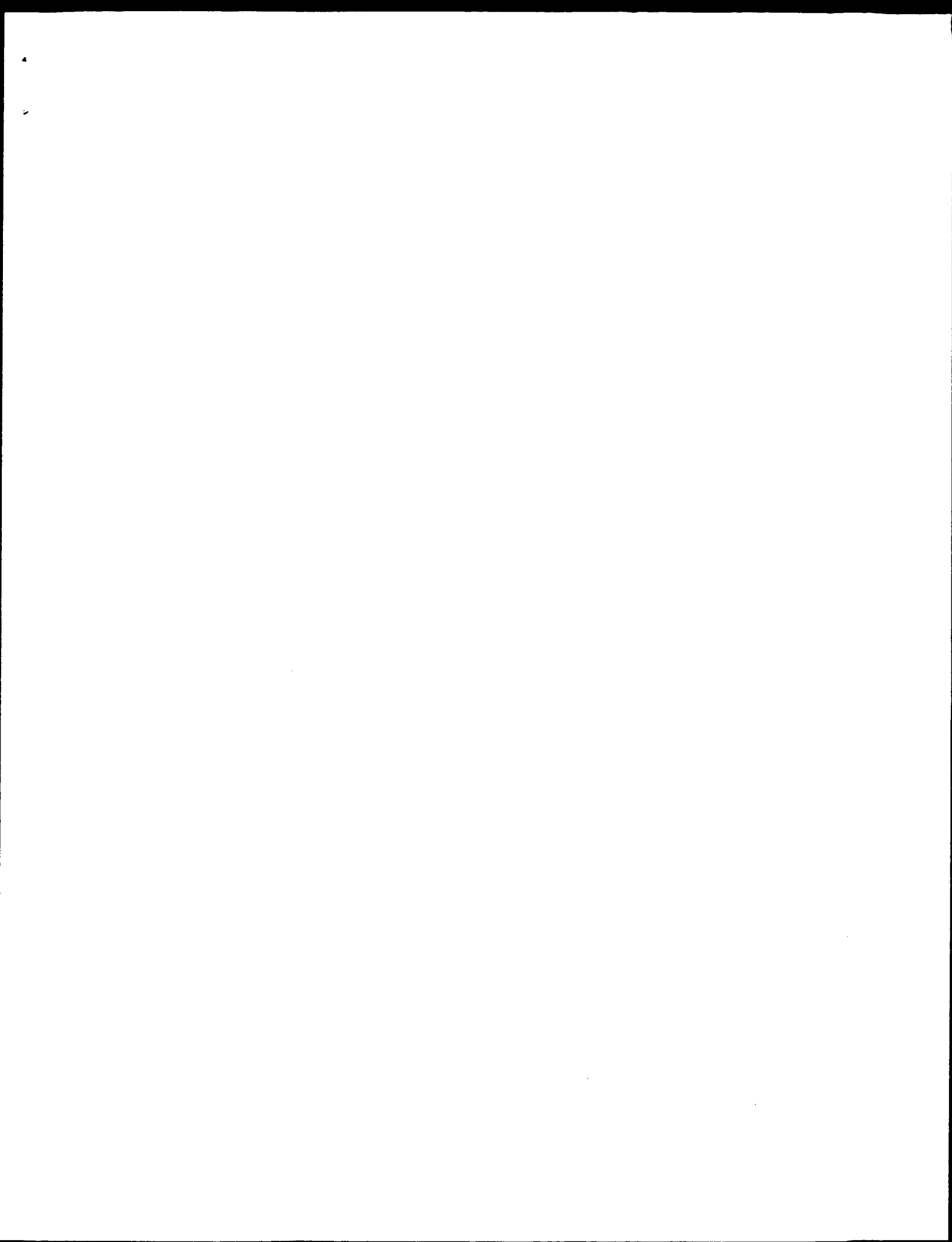
PS Disclosure: Column 97-126; 109pp: English.

XX This sequence is a fragment of the Kaposi's sarcoma-associated
CC herpesvirus (KSHV) LUR (long unique region). This fragment contains
CC coding regions for ORF20, ORF21 which encodes thymidine kinase, ORF22
CC which encodes glycoprotein H, ORF23, ORF24, ORF25 which encodes the
CC major capsid protein, ORF26 which encodes capsid protein II, ORF27,
CC ORF28, ORF29b which encodes packaging protein II, ORF30, ORF31, ORF32,
CC ORF33, ORF39a which encodes packaging protein I, ORF34, ORF35, ORF36
CC which encodes viral protein kinase, ORF37 which encodes alkaline
CC exonuclease, ORF38, ORF39 which encodes glycoprotein M, ORF40 which
CC encodes helicase subunit I, ORF41 which encodes helicase primase
CC subunit II, ORF42, ORF43 which encodes capsid protein III, ORF44 which
CC encodes helicase primase subunit III, ORF45 which encodes the virion
CC assembly protein, ORF46 which encodes uracil DNA glycosylase (UDG),
CC ORF47 which encodes glycoprotein L. KSHV is a new human Herpesvirus
CC (HHV8) believed to cause Kaposi's sarcoma (KS) which is the most common
CC form of neoplasm occurring in persons with acquired immune deficiency
CC syndrome (AIDS). The DHFR protein is useful for vaccination, prophylaxis,
CC diagnosis and treatment of a subject with Kaposi's sarcoma and for
CC detecting expression of a DNA virus associated with Kaposi's sarcoma in a
CC cell.

XX Sequence 35100 BP; 8016 A; 9231 C; 9477 G; 8376 T; 0 other:
SQ

Query Match 56.0%; Score 14; DB 20; Length 35100;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 tgcacgcacacat 19
|||||
DB 2058 TGCATCGCCAGCAT 2045

Search completed: December 26, 2001, 14:16:47
Job time: 6999 sec



Thu Dec 27 16:12:35 2001

us-09-396-196f-6.olig.rml

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:53:23 : Search time 143.1 Seconds
(without alignments)
39.566 Million cell updates/sec

Title: US-09-396-196f-6

Perfect score: 25

Sequence: 1 gcagtgatgcgcagcattcgat 25

Scoring table: OLIGO-MNC
Gapop 60.0, Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size: 9

Total number of hits satisfying chosen parameters: 10660

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database: Issued_Patents_NA:
1: /cgn2_6/prodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCRTUS.COMB.seq:*
6: /cgn2_6/prodata/2/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appl
2	25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appl
3	25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appl
4	25	100.0	5872	3	US-08-411-768B-6	Sequence 45, Appl
5	14	56.0	718	4	US-09-142-078-45	Sequence 45, Appl
6	14	56.0	718	4	US-09-357-141-45	Sequence 16, Appl
7	14	56.0	2193	1	US-08-420-235B-16	Sequence 16, Appl
8	14	56.0	2193	1	US-08-793-624-16	Sequence 16, Appl
9	14	56.0	2193	5	PCT-US95-10194-16	Sequence 16, Appl
10	14	56.0	2232	5	PCT-US96-05320A-113	Sequence 113, Appl
11	14	56.0	2973	2	US-08-343-101A-5	Sequence 5, Appl
12	14	56.0	2973	3	US-09-183-688-5	Sequence 5, Appl
13	14	56.0	4223	4	US-09-541-782-5	Sequence 5, Appl
14	14	56.0	5319	1	US-08-169-927-1	Sequence 1, Appl
15	14	56.0	20710	3	US-08-420-235B-1	Sequence 1, Appl
16	14	56.0	20710	3	US-08-793-624-1	Sequence 1, Appl
17	14	56.0	20710	5	PCT-US95-10194-1	Sequence 1, Appl
18	14	56.0	35100	2	US-08-770-379-18	Sequence 18, Appl
19	14	56.0	35100	2	US-08-757-669A-18	Sequence 726, Appl
20	13	52.0	304	4	US-09-060-756-726	Sequence 570, Appl
21	13	52.0	304	4	US-09-060-756-726	Sequence 570, Appl
22	13	52.0	447	4	US-08-836-075A-51	Sequence 51, Appl
23	13	52.0	1460	2	US-08-933-750C-60	Sequence 60, Appl
24	13	52.0	1460	3	US-09-234-613-60	Sequence 60, Appl
25	13	52.0	1624	2	US-07-862-588B-6	Sequence 6, Appl
26	13	52.0	1775	2	US-07-862-588B-6	Sequence 5, Appl
27	13	52.0	2290	1	US-08-584-226-1	Sequence 1, Appl

ALIGNMENTS

```
RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 7.8e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcacgcacacatttcgat 25
   |||||||||||||||||||
DB 84 GCAGGTGCATCCGCACATTTCGAT 108

RESULT 2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
```

```
ADDRESS: No. 5869719artis Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 7.8e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcacgcacacatttcgat 25
   |||||||||||||||||||
DB 84 GCAGGTGCATCCGCACATTTCGAT 108

RESULT 3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Owen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; OF PRODUCING BIOTIN
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect
; SOFTWARE: Version 5.1
; CURRENT APPLICATION DATA:
```

```

APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
SPRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pB030A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=117
OTHER INFORMATION: /product= "biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biob"
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
OTHER INFORMATION: /codon_start= 2295
OTHER INFORMATION: /function= "Involved in pimeoyl-CoA synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "bioc"
OTHER INFORMATION: /number= 3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3750
OTHER INFORMATION: /pc_number= 2.6.1.62
OTHER INFORMATION: /product= "DAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioA"
OTHER INFORMATION: /number= 5
OTHER INFORMATION: /standard_name=
OTHER INFORMATION: "S-Adenosyl-L-methionine:8-amino-7-oxononanoate"
OTHER INFORMATION: "aminotransf."
FEATURE:
NAME/KEY: CDS
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 5098
OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ORF1"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: -10-signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "promoter ptac"
FEATURE:
NAME/KEY: -35-signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name= "promoter ptac"

```

```

FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name= "bioA RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5068..5100
OTHER INFORMATION: /standard_name= "ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name= "rho-independent"
OTHER INFORMATION: transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "promoter ptac"
OTHER INFORMATION: /evidence= EXPERIMENTAL
PUBLICATION NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match          100.0%; Score 25; Length 5872;
Best Local Similarity 100.0%; Pred. No. 8e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 gcagtgacatgcgcagcattcgat 25
         |||
Db       200 GCAGTGCATCGCAGCATTGCGAT 224

RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect
; SOFTWARE: Version 5.1

```

```

1  GENERAL INFORMATION:
2  APPLICANT: McCabe, R. Tyler
3  APPLICANT: Zhou, Li-Ming
4  APPLICANT: Layer, Richard T.
5  TITLE OF INVENTION: Use of Conantokins
6  NUMBER OF SEQUENCES: 71
7  CORRESPONDENCE ADDRESS:
8  ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
9  STREET: 555 Thirteenth Street, N.W., Suite 701-E
10 CITY: Washington
11 STATE: D.C.
12 COUNTRY: USA
13 ZIP: 20004
14
15 COMPUTER READABLE FORM:
16
17 MEDIUM TYPE: Floppy disk
18 COMPUTER: IBM PC compatible
19 OPERATING SYSTEM: PC-DOS/MS-DOS
20 SOFTWARE: Patentin Release #1.0, Version #1.30
21 CURRENT APPLICATION DATA:
22 APPLICATION NUMBER: US/09/142,078
23 FILING DATE: 10-FEB-1999
24
25 PRIORITY INFORMATION:
26 APPLICATION NUMBER: WO 99/12652
27 FILING DATE: 21-JUL-1997
28
29 PRIORITY INFORMATION:
30 APPLICATION NUMBER: US 08/762,377
31 FILING DATE: 06-DEC-1996
32
33 PRIORITY INFORMATION:
34 APPLICATION NUMBER: US 08/684,750
35 FILING DATE: 22-JUL-1996
36 ATTORNEY/AGENT INFORMATION:
37 NAME: Ihnen, Jeffrey L.
38 REGISTRATION NUMBER: 28,957
39 REFERENCE/DOCKET NUMBER: 2314-135.A
40 TELECOMMUNICATION INFORMATION:
41 TELEPHONE: 202-783-6040
42 TELEFAX: 202-783-6031
43
44 INFORMATION FOR SEQ ID NO: 45:
45 SEQUENCE CHARACTERISTICS:
46 LENGTH: 718 base pairs
47 TYPE: nucleic acid
48 STRANDEDNESS: single
49 TOPOLOGY: linear
50 MOLECULE TYPE: DNA (cDNA)
51 ORIGINAL SOURCE:
52 ORGANISM: Conus geographus
53 FEATURE:
54 NAME/KEY: COS
55 LOCATION: 110..409
56 US-09-142-078-45
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
9
```

CURRENT APPLICATION NUMBER: US/09/357,141
CURRENT FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: US 09/283,277
PRIOR FILING DATE: 1999-04-01
PRIOR APPLICATION NUMBER: US 09/142,078
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: WO US97/12652
PRIOR FILING DATE: 1997-07-21
PRIOR APPLICATION NUMBER: US 08/762,377
PRIOR FILING DATE: 1996-12-06
PRIOR APPLICATION NUMBER: US 08/684,750
PRIOR FILING DATE: 1996-07-22
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 45
LENGTH: 718
TYPE: DNA
ORGANISM: Conus geographus
FEATURE:
NAME/KEY: CDS
LOCATION: (110)..(409)
US-09-357-141-45

Query Match 56.0%; Score 14; DB 4; Length 718;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cagatgcagccca 15
|||||
Db 118 CAGGTCATGCCCA 105

RESULT 7
US-08-420-235B-16/C
Sequence 16, Application US/08420235B
Patent No. 5801042
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/420,235B
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2193 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2193
OTHER INFORMATION:
US-08-420-235B-16

Query Match 56.0%; Score 14; DB 1; Length 2193;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcagccagcat 19
|||||
Db 46 TGCATGCCAGCAT 33

RESULT 8
US-08-793-624-16/C
Sequence 16, Application US/08793624C
Patent No. 6150093
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And
TITLE OF INVENTION: Uses Thereof
FILE REFERENCE: 45185-C-PCT-US/JPM
CURRENT APPLICATION NUMBER: US/08/793,624C
CURRENT FILING DATE: 1997-02-18
NUMBER OF SEQ ID NOS: 58
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 16
LENGTH: 2193
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-08-793-624-16

Query Match 56.0%; Score 14; DB 3; Length 2193;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcagccagcat 19
|||||
Db 46 TGCATGCCAGCAT 33

RESULT 9
PCT-US95-10194-16/C
Sequence 16, Application PC/TUS9510194
GENERAL INFORMATION:
APPLICANT: The Trustees of Columbia University in the City of New York
APPLICANT: City
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10194
FILING DATE:
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPM/MS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ. ID NO.: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2193 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2193
OTHER INFORMATION:
PCT-US95-10194-16

Query Match
Best Local Similarity 56.0%; Score 14; DB 5; Length 2193;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 6 tgcacgcacgcat 19
Db 46 tgcacgcacgcat 33

RESULT 10
PCT-US96-05320A-113/c
Sequence 113, Application PC/TUS9605320A
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences
APPLICANT: 9410 Key West Avenue
APPLICANT: Rockville, MD 20850
APPLICANT: United States of America
APPLICANT: Johns Hopkins University
APPLICANT: 720 Rutland Avenue
APPLICANT: Baltimore, MD 21205
APPLICANT: United States of America
APPLICANT: Mark D. Adams
APPLICANT: Owen White
APPLICANT: Hamilton O. Smith
TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20003-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/05320A
FILING DATE: April 22, 1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: June 7, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: June 7, 1995
ATTORNEY/AGENT INFORMATION:

NAME: Eric K. Steffe
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.014PC01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ. ID NO.: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 2232 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
PCT-US96-05320A-113

Query Match
Best Local Similarity 56.0%; Score 14; DB 5; Length 2232;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 9 atcgccacatttc 22
Db 636 atcgccacatttc 623

RESULT 11
US-08-343-101A-5/c
Sequence 5, Application US/08343101A
Patent No. 5830759
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,101A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: White Esq., John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0526
INFORMATION FOR SEQ. ID NO.: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2973 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
US-08-343-101A-5

Query Match
Best Local Similarity 56.0%; Score 14; DB 2; Length 2973;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacacat 19
 |||
 DB 2182 TGCATGCCAGCAT 2169

RESULT 12
 US-09-183-688-5/c
 ; Sequence 5, Application US/09183688
 ; Patent No. 6093550

GENERAL INFORMATION:
 APPLICANT: Chang, Yuan
 APPLICANT: Moore, Patrick S.
 TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma
 TITLE OF INVENTION: Virus Sequences And Uses Thereof
 NUMBER OF SEQUENCES: 22
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Cooper & Dunham
 STREET: 1185 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.24
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/183,688

CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/343,101
 FILING DATE:

ATTORNEY/AGENT INFORMATION:
 NAME: White Esq., John P.
 REGISTRATION NUMBER: 28,678
 REFERENCE/DOCKET NUMBER: 45185-A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-278-0400
 TELEFAX: 212-391-0526

INFORMATION FOR SEQ. ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2973 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: Linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: N
 ANTI-SENSE: N
 US-09-183-688-5

Query Match 56.0%; Score 14; DB 3; Length 2973;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacacat 19
 |||
 DB 2182 TGCATGCCAGCAT 2169

RESULT 13
 US-09-541-782-5/c
 ; Sequence 5, Application US/09541782
 ; Patent No. 6284480

GENERAL INFORMATION:
 APPLICANT: Nislow, Corey
 APPLICANT: Sakowicz, Roman
 APPLICANT: Beraud, Christophe
 TITLE OF INVENTION: Antifungal Assay
 FILE REFERENCE: 1015
 CURRENT APPLICATION NUMBER: US/09/541,782

CURRENT FILING DATE: 2000-04-03
 NUMBER OF SEQ. ID NOS: 10
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ. ID NO 5
 LENGTH: 4223
 TYPE: DNA
 ORGANISM: S.pombe
 US-09-541-782-5

Query Match 56.0%; Score 14; DB 4; Length 4223;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgcacacatt 21
 |||
 DB 1559 CATGCCAGCATTT 1546

RESULT 14
 US-08-169-927-1/c
 ; Sequence 1, Application US/08169927
 ; Patent No. 5783441

GENERAL INFORMATION:
 APPLICANT: Carl, Mitchell
 APPLICANT: Dobson, Michael E.
 APPLICANT: Ching, Wei Mei
 APPLICANT: Dasch, Gregory A.
 TITLE OF INVENTION: Gene and Protein Applicable to the
 TITLE OF INVENTION: Preparation of Vaccines for Rickettsia prowazekii and
 TITLE OF INVENTION: Rickettsia typhi and the Detection of Both
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Counsel, Naval Medical R & D Command
 STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
 CITY: Bethesda
 STATE: MD
 COUNTRY: USA
 ZIP: 20889-5606

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/169,927
 FILING DATE:

CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/742,128
 FILING DATE: 08/09/91
 ATTORNEY/AGENT INFORMATION:
 NAME: Spevack, A. David
 REGISTRATION NUMBER: 24,743
 REFERENCE/DOCKET NUMBER: 75,976
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 295-6759
 TELEFAX: (301) 295-1022
 INFORMATION FOR SEQ. ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 5319 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: circular
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: internal
 ORIGINAL SOURCE: Rickettsia prowazekii
 ORGANISM: Rickettsia prowazekii
 STRAIN: Breinl
 FEATURE:
 NAME/KEY: -35_signal

LOCATION: 340..345
FEATURE: -10_signal
NAME/KEY: 363..368
LOCATION: 363..368
FEATURE: CDS
NAME/KEY: 391..5226
LOCATION: 391..5226
FEATURE: RBS
NAME/KEY: 379..386
LOCATION: 379..386
FEATURE: stem_loop
NAME/KEY: 5270..5306
LOCATION: 5270..5306
PUBLICATION INFORMATION:
AUTHORS: Carl, M.
AUTHORS: Dobson, M. E.
AUTHORS: Ching, W. M.
AUTHORS: Dasch, G. A.
TITLE: Characterization of the gene encoding the
TITLE: protective S-layer protein of Rickettsia
TITLE: prowazekii; presence of a truncated identical
TITLE: homolog in rickettsia typhi
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
DATE: 1990
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 5319
US-08-169-927-1

Query Match 56.0%; Score 14; DB 1; Length 5319;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 tcatgccagcat 21
|||||
Db 2221 CATGCCAGCATTT 2208

RESULT 15
US-08-420-235B-1/c
Sequence 1, Application US/08420235B
Patent No. 5801042
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/420,235B
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 20710 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
US-08-420-235B-1

Query Match 56.0%; Score 14; DB 1; Length 20710;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcagcat 19
|||||
Db 2138 TGCATGCCAGCAT 2125

RESULT 16
US-08-793-624-1/c
Sequence 1, Application US/08793624C
Patent No. 6150093
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And
TITLE OF INVENTION: Uses Thereof
FILE REFERENCE: 45185-C-PCT-US/JPM
CURRENT APPLICATION NUMBER: US/08/793,624C
CURRENT FILING DATE: 1997-02-18
NUMBER OF SEQ ID NOS: 58
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 1
LENGTH: 20710
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-08-793-624-1

Query Match 56.0%; Score 14; DB 3; Length 20710;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcagcat 19
|||||
Db 2138 TGCATGCCAGCAT 2125

RESULT 17
PCT-US95-10194-1/c
Sequence 1, Application PC/TUS9510194
GENERAL INFORMATION:
APPLICANT: The Trustees of Columbia University in the City of New York
APPLICANT: City
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10194
FILING DATE:
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPW/MS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 20710 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
PCT-US95-10194-1

Query Match 56.0%; Score 14; DB 5; Length 20710;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacgat 19
|||||
DB 2138 TGCATGCCGACGAT 2125

RESULT 18
US-08-770-379-18/C
Sequence 18, Application US/08770379
Patent No. 5849564
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,379
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 35100 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-770-379-18

Query Match 56.0%; Score 14; DB 2; Length 35100;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacgat 19
|||||
DB 2058 TGCATGCCGACGAT 2045

RESULT 19
US-08-757-669A-18/C
Sequence 18, Application US/08757669A
Patent No. 6183751
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/757,669A
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-F
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 35100 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-757-669A-18

Query Match 56.0%; Score 14; DB 4; Length 35100;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacgat 19
|||||
DB 2058 TGCATGCCGACGAT 2045

RESULT 20
US-09-060-756-726
Sequence 726, Application US/09060756
Patent No. 6183957
GENERAL INFORMATION:
APPLICANT: Cole, Stewart
APPLICANT: Buchrieser-Brosch, Roland
APPLICANT: Gordon, Stephen

APPLICANT: Billault, Alain
TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
FILE REFERENCE: 3495-0169
CURRENT APPLICATION NUMBER: US/09/060,756
CURRENT FILING DATE: 1998-04-16
NUMBER OF SEQ ID NOS: 743
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 726
LENGTH: 304
TYPE: DNA
ORGANISM: Mycobacterium bovis
FEATURE:
NAME/KEY: unsure
LOCATION: (various positions within the sequence)
OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-060-756-726

Query Match
Best Local Similarity 52.0%; Score 13; DB 4; Length 304;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 gtgcacgcgcgcgc 17
DB 136 gtgcacgcgcgcgc 148

RESULT 21
US-09-060-756-570
Sequence 570, Application US/09060756
Patent No. 6183957
GENERAL INFORMATION:
APPLICANT: Cole, Stewart
APPLICANT: Buchrieser-Brosch, Roland
APPLICANT: Gordon, Stephen
APPLICANT: Billault, Alain
TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
FILE REFERENCE: 3495-0169
CURRENT APPLICATION NUMBER: US/09/060,756
CURRENT FILING DATE: 1998-04-16
NUMBER OF SEQ ID NOS: 743
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 570
LENGTH: 343
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
US-09-060-756-570

Query Match
Best Local Similarity 52.0%; Score 13; DB 4; Length 343;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 gtgcacgcgcgcgc 17
DB 135 gtgcacgcgcgcgc 147

RESULT 22
US-08-836-075A-51
Sequence 51, Application US/08836075A
Patent No. 6180768
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
APPLICANT: STUYVER, LIEVEN
TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
NUMBER OF SEQUENCES: 207

CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,075A
FILING DATE: 21 Apr 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/04155
FILING DATE: 23 Oct 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHEICAL: NO
ANTI-SENSE: NO
US-08-836-075A-51

Query Match
Best Local Similarity 52.0%; Score 13; DB 4; Length 447;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgcgcgcgcgc 20
DB 406 CATGCgcgcgcgc 418

RESULT 23
US-08-933-750C-60
Sequence 60, Application US/08933750C
Patent No. 5932442
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purni
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933.750C
FILING DATE: September 23, 1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 1460 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: TLYMNOR01
CLONE: 140516
US-08-933-750C-60

Query Match 52.0%; Score 13; DB 2; Length 1460;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 13 ccagcattcgat 25
|||||
Db 722 CCAGCATTTCGAT 734

RESULT 24
US-09-234-613-60
Sequence 60; Application US/09224613
Patent No. 6132873
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/234.613
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/933.750
FILING DATE: September 23, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 1460 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: TLYMNOR01
CLONE: 140516
US-09-234-613-60

Query Match 52.0%; Score 13; DB 3; Length 1460;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 13 ccagcattcgat 25
|||||
Db 722 CCAGCATTTCGAT 734

RESULT 25
US-07-862-588B-6
Sequence 6; Application US/07862588B
Patent No. 5916796
GENERAL INFORMATION:
APPLICANT: Joergensen, Per Linnae
APPLICANT: Sch Jein, Martin
APPLICANT: Hansen, Christian
TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
STREET: 405 Lexington Avenue, 62nd floor
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10017
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/862,588B
FILING DATE: 19920727
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 164/90
FILING DATE: 19-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK91/00013
FILING DATE: 18-JAN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Zelson, Steve T. / Lambiris, Elias J.
REGISTRATION NUMBER: 30,335 / 33,728
REFERENCE/DOCKET NUMBER: 3425.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1624 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Bacillus lautus
STRAIN: NCIMB 40250
FEATURE:
NAME/KEY: CDS
LOCATION: 30..1607
OTHER INFORMATION:
US-07-862-588B-6

Query Match 52.0%; Score 13; DB 2; Length 1624;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 caggtgcacgcc 14
|||||
DB 934 CAGGTGCATCGCC 946

RESULT 26
US-07-862-588B-5
Sequence 5, Application US/07862588B
Patent No. 5916796
GENERAL INFORMATION:
APPLICANT: Joergensen, Per Linna
APPLICANT: Sch lein, Martin
APPLICANT: Hansen, Christian
TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5916796 No. 5916796disk of No. 5916796th America, Inc.
STREET: 405 Lexington Avenue, 62nd floor
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10017
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/862,588B
FILING DATE: 19920727
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 164/90
FILING DATE: 19-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK91/00013
FILING DATE: 18-JAN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Zelson, Steve T. / Lambiris, Elias J.
REGISTRATION NUMBER: 30,335 / 33,728
REFERENCE/DOCKET NUMBER: 3425.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1775 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Bacillus lautus
STRAIN: NCIMB 40250
FEATURE:

NAME/KEY: CDS
LOCATION: 30..(1625,1775)
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /partial
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /transl_except= (pos: 1446 .. 1458, aa:
US-07-862-588B-5

Query Match 52.0%; Score 13; DB 2; Length 1775;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 caggtgcacgcc 14
|||||
DB 934 CAGGTGCATCGCC 946

RESULT 27
US-08-584-226-1/c
Sequence 1, Application US/08584226
Patent No. 5798240
GENERAL INFORMATION:
APPLICANT: Martinis, Susan A.
APPLICANT: Sassanfar, Mandana
APPLICANT: Kim, Sungmoon
APPLICANT: Lee, Sang Ho
APPLICANT: Schimmel, Paul R.
TITLE OF INVENTION: RECOMBINANT MYCOBACTERIAL METHIONYL-tRNA
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173-4799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/584,226
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/305,766
FILING DATE: 13-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: CP194-052
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2290 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1563
US-08-584-226-1

Query Match 52.0%; Score 13; DB 1; Length 2290;
Best Local Similarity 100.0%; Pred. No. 48;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 caggtgcatgcc 14
|||||
Db 1233 CAGGTGCATGCC 1221

RESULT 28

US-09-134-566-15
; Sequence 15, Application US/09134566
; Patent No. 5998147
; GENERAL INFORMATION:
; APPLICANT: Petit, Christine
; APPLICANT: Denoyelle-Gryson, Francoise
; APPLICANT: Weil, Dominique
; APPLICANT: Marlin-Duvernois, Sandrine
; APPLICANT: Guesdon, Jean-Luc
; TITLE OF INVENTION: MUTATED POLYNUCLEOTIDE CORRESPONDING TO A MUTATION
; TITLE OF INVENTION: RESPONSIBLE FOR PRELINGUAL NON-SYNDROMIC DEAFNESS IN
; TITLE OF INVENTION: THE CONNEXIN 26 GENE AND METHOD OF DETECTING THIS
; TITLE OF INVENTION: HEREDITARY DEFECT
; FILE REFERENCE: 03495-0175
; CURRENT APPLICATION NUMBER: US/09/134,566
; CURRENT FILING DATE: 1998-08-14
; EARLIER APPLICATION NUMBER: 60/055,863
; EARLIER FILING DATE: 1998-08-15
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 2312
; TYPE: DNA
; ORGANISM: human connexin
US-09-134-566-15

Query Match 52.0%; Score 13; DB 2; Length 2312;

Best Local Similarity 100.0%; Pred. No. 48;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 tcgccagcatcc 22
|||||
Db 960 tcgccagcatcc 972

RESULT 29

US-09-134-566-8
; Sequence 8, Application US/09134566
; Patent No. 5998147
; GENERAL INFORMATION:
; APPLICANT: Petit, Christine
; APPLICANT: Denoyelle-Gryson, Francoise
; APPLICANT: Weil, Dominique
; APPLICANT: Marlin-Duvernois, Sandrine
; APPLICANT: Guesdon, Jean-Luc
; TITLE OF INVENTION: MUTATED POLYNUCLEOTIDE CORRESPONDING TO A MUTATION
; TITLE OF INVENTION: RESPONSIBLE FOR PRELINGUAL NON-SYNDROMIC DEAFNESS IN
; TITLE OF INVENTION: THE CONNEXIN 26 GENE AND METHOD OF DETECTING THIS
; FILE REFERENCE: 03495-0175
; CURRENT APPLICATION NUMBER: US/09/134,566
; CURRENT FILING DATE: 1998-08-14
; EARLIER APPLICATION NUMBER: 60/055,863
; EARLIER FILING DATE: 1998-08-15
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 2314
; TYPE: DNA
; ORGANISM: human connexin
US-09-134-566-8

Query Match 52.0%; Score 13; DB 2; Length 2314;

Best Local Similarity 100.0%; Pred. No. 48;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 tcgccagcatcc 22
|||||
Db 962 tcgccagcatcc 974

RESULT 30

US-08-759-444-2/c
; Sequence 2, Application US/08759444
; Patent No. 5824309
; GENERAL INFORMATION:
; APPLICANT: Dassarma, Shiladitya
; APPLICANT: Morshed, Fazela
; APPLICANT: Stuart, Elizabeth
; APPLICANT: Black, Samuel
; TITLE OF INVENTION: RECOMBINANT GAS VESICLES AND USES THEREOF
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/759,444
; FILING DATE: 05-DEC-1996
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 60/008,200
; FILING DATE: 05-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Melkielejohn, Ph.D., Anita L
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07880/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8878 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-759-444-2

Query Match 52.0%; Score 13; DB 1; Length 8878;

Best Local Similarity 100.0%; Pred. No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 gtgcacgcagc 17
|||||
Db 7684 GTGCATGCCAGC 7672

RESULT 31

US-08-680-897-1/c
; Sequence 1, Application US/08680897
; Patent No. 6008051
; GENERAL INFORMATION:
; APPLICANT: Dassarma, Shiladitya
; APPLICANT: Halladay, John
; APPLICANT: Ng, Wai-lap
; TITLE OF INVENTION: RECOMBINANT VECTOR AND PROCESS FOR CELL
; TITLE OF INVENTION: FLUTATION

Query Match 52.0%; Score 13; DB 2; Length 2314;

```

RESULT 32
US-09-103-840A-2
: Sequence 2, Application US/09103840A
: Patent No. 6294328
:
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
:
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103,840A
: CURRENT FILING DATE: 1998-06-24
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.1
:
: SEQ ID NO 2
:
: LENGTH: 4403765
:
: TYPE: DNA
:
: ORGANISM: Mycobacterium tuberculosis
:
: FEATURE:
:
: OTHER INFORMATION: CDC 1551
:
: OTHER INFORMATION: "n" bases at various positions throughout the sequence
: OTHER INFORMATION: represent a, t, c or g

```

```

1 RESULT 34
2 US-09-103-840A-1/C
3 ; Sequence 1, Application US/09103840A
4 ; Patent No. 6294328
5 ; GENERAL INFORMATION:
6 ; APPLICANT: FLEISCHMAN, Robert D.
7 ; APPLICANT: WHITE, Owen R.
8 ; APPLICANT: FRASER, Claire M.
9 ; APPLICANT: VENTER, John C.
10 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
11 ; FILE REFERENCE: 24366-20007 00
12 ; CURRENT APPLICATION NUMBER: US/09/103,840A
13 ; NUMBER OF SEQ. ID NOS: 2
14 ; SOFTWARE: PatentIn Ver. 2.1
15 ; SEQ. ID NO. 1
16 ; LENGTH: 4411529
17 ; TYPE: DNA
18 ; ORGANISM: Mycobacterium tuberculosis
19 ; OTHER INFORMATION: H37Rv
20 US-09-103-840A-1

```

```
Query Match      52.0%; Score 13; DB 4; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 1 gcaggtgcagccca 15
|||||
Db 2148914 GCAGGTGCATCGCCA 2148900

RESULT 35
US-08-851-843A-92/C
Sequence 92, Application US/08851843A
Patent No. 6093809
GENERAL INFORMATION:
APPLICANT: Cec'h, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morlin, Gregg B.
APPLICANT: Harley, Calvin B.
APPLICANT: Andrews, William H.
TITLE OF INVENTION: No. 6093809el Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,843A
FILING DATE: 06-MAY-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-0029300S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 92:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-851-843A-92

Query Match 48.0%; Score 12; DB 3; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgcagcat 19
|||||
Db 19 CATGCCAGCAT 8

RESULT 36
US-08-974-549A-384/C
Sequence 384, Application US/08974549A
Patent No. 6166178
GENERAL INFORMATION:
APPLICANT: Cec'h, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morlin, Gregg B.
APPLICANT: Harley, Calvin B.
APPLICANT: Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 727
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-NOV-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-0026100S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 384:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: -
LOCATION: 1..19
OTHER INFORMATION: /note="TCTP.6 primer"
US-08-974-549A-384

Query Match 48.0%; Score 12; DB 4; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catcgccagcat 19
|||||
DB 19 CATGCCAGCAT 8

RESULT 37
US-08-854-050-92/c
Sequence 92, Application US/08854050
Patent No. 6261836
GENERAL INFORMATION:
APPLICANT: Cecch, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin
APPLICANT: Andrews, William H.
TITLE OF INVENTION: No. 6261836el Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/854,050
FILING DATE: 09-MAY-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 92:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-854-050-92

Query Match 48.0%; Score 12; DB 4; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catcgccagcat 19
|||||
DB 19 CATGCCAGCAT 8

RESULT 38
US-08-373-124A-1692/c
Sequence 1692, Application US/08373124A
Patent No. 5646042
GENERAL INFORMATION:
APPLICANT: Stinchcomb, Dan T.
APPLICANT: Draper, Kenneth
APPLICANT: McSwigen, James
APPLICANT: Jarvis, Thale
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
TITLE OF INVENTION: TREATMENT OF RESTENOSIS AND
NUMBER OF SEQUENCES: 2627
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Suite 4700
STATE: Los Angeles
COUNTRY: California
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/373,124A
FILING DATE: January 13, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/245,466
FILING DATE: May 18, 1994
APPLICATION NUMBER: 08/192,943
FILING DATE: February 7, 1994
APPLICATION NUMBER: 07/987,132
FILING DATE: December 7, 1992
APPLICATION NUMBER: 07/936,422
FILING DATE: August 26, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Wardburg, Richard
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 209/035
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1692:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-373-124A-1692

Query Match 48.0%; Score 12; DB 1; Length 38;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 gccagcatctcg 23
|||||
DB 38 GCCAGCATTTTCG 27

RESULT 39
US-08-435-628-1692/c
Sequence 1692, Application US/08435628
Patent No. 5817796

GENERAL INFORMATION:

APPLICANT: Stinchcomb, Dan T.
APPLICANT: Draper, Kenneth
APPLICANT: McSwigen, James
APPLICANT: Jarvis, Thale
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
TITLE OF INVENTION: TREATMENT OF RESTENOSIS AND
TITLE OF INVENTION: CANCER USING RIBOZYMES
NUMBER OF SEQUENCES: 2627
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 MB
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/435,628
FILING DATE: 05-MAY-1995

CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/373,124
FILING DATE: January 13, 1995
APPLICATION NUMBER: 08/245,466
FILING DATE: May 18, 1994
APPLICATION NUMBER: 08/192,943
FILING DATE: February 7, 1994
APPLICATION NUMBER: 07/987,132
FILING DATE: December 7, 1992
APPLICATION NUMBER: 07/936,422
FILING DATE: August 26, 1992

ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 209/035
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 488-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1692:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-435-628-1692

Query Match 48.0%; Score 12; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 12 gccagcatctcg 23
|||||

DB 38 GCCAGCATTTTCG 27

RESULT 40
US-09-023-228B-59

Sequence 59, Application US/09023228B
Patent No. 6140490

GENERAL INFORMATION:

APPLICANT: HIESECKER, GREGORY
APPLICANT: GOLD, LARRY
TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS OF
TITLE OF INVENTION: COMPLEMENT SYSTEM PROTEINS
NUMBER OF SEQUENCES: 157
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Place #200
CITY: Denver
STATE: Colorado
COUNTRY: US

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,228B
FILING DATE: 12-FEBRUARY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/01739
FILING DATE: 30 JAN 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/595,335
FILING DATE: 1 FEB 1996

ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson, Esq.
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX50/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433

INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 61 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA
FEATURE:
OTHER INFORMATION: All pyrimidines are 2'-F modified
US-09-023-228B-59

Query Match 48.0%; Score 12; DB 3; Length 61;
Best Local Similarity 66.7%; Pred. No. 1.7e+02;
Matches 8; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 14 gccagcatcgat 25
|||||
DB 23 CAGCAUUCGACAU 34

RESULT 41

US-08-488-402A-16/c
Sequence 16, Application US/08488402A
Patent No. 5837456

GENERAL INFORMATION:

APPLICANT: GOLD ET AL.
TITLE OF INVENTION: HIGH AFFINITY OLIGONUCLEOTIDE LIGANDS TO
TITLE OF INVENTION: CHORIONIC GONADOTROPIN HORMONE AND RELATED GLYCOPROTEIN
TITLE OF INVENTION: HORMONES
NUMBER OF SEQUENCES: 160
CORRESPONDENCE ADDRESS:

ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MG
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,402A
FILING DATE: 07 JUNE 1995

CLASSIFICATION: 536
APPLICATION NUMBER: 07/536,428
FILING DATE: 11-JUNE-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/714,131
FILING DATE: 10-JUNE-1991

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/931,473
FILING DATE: 17-AUGUST-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/964,624
FILING DATE: 21-OCTOBER-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/177,991
FILING DATE: 8-SEPTEMBER-1993

ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215

REFERENCE/DOCKET NUMBER: NEX36-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333

TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:

LENGTH: 69 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
FEATURE:
OTHER INFORMATION: All C's are 2'-NH2 modified

FEATURE:
OTHER INFORMATION: All U's are 2'-NH2 modified
US-08-488-402A-16

Query Match 48.0%; Score 12; DB 2; Length 69;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 aggtgcatcgcc 14
DB 25 AGGTGATCGCC 14

RESULT 42

US-08-484-552A-16/C
Sequence 16, Application US/08484552A
GENERAL INFORMATION:
PATENT NO. 5849890

APPLICANT: GOLD, LARRY
APPLICANT: JAYASNA, SURENDA
APPLICANT: NIEMULANDT, DAN

APPLICANT: DAVIS, KEN
TITLE OF INVENTION: HIGH AFFINITY OLIGONUCLEOTIDE LIGANDS TO CHORIONIC
NUMBER OF INVENTION: GONADOTROPIN HORMONE AND RELATED GLYCOPROTEIN HORMONES

NUMBER OF SEQUENCES: 160
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.

STREET: 8400 E. Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MG
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 6.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,552A
FILING DATE: 07 JUNE 1995

CLASSIFICATION: 536
APPLICATION NUMBER: 07/536,428
FILING DATE: 11-JUNE-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/714,131
FILING DATE: 10-JUNE-1991

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/931,473
FILING DATE: 17-AUGUST-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/964,624
FILING DATE: 21-OCTOBER-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/177,991
FILING DATE: 8-SEPTEMBER-1993

ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215

REFERENCE/DOCKET NUMBER: NEX36-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333

TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:

LENGTH: 69 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
FEATURE:
OTHER INFORMATION: All C's are 2'-NH2 modified

FEATURE:
OTHER INFORMATION: All U's are 2'-NH2 modified
US-08-484-552A-16

Query Match 48.0%; Score 12; DB 2; Length 69;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 aggtgcatcgcc 14
DB 25 AGGTGATCGCC 14

RESULT 43

PCT-US96-09472-16/C
Sequence 16, Application PC/TUS9609472
GENERAL INFORMATION:
PATENT NO. 5849890

APPLICANT: GOLD ET AL.
TITLE OF INVENTION: HIGH AFFINITY OLIGONUCLEOTIDE
LIGANDS TO CHORIONIC GONADOTROPIN HORMONE AND RELATED

NUMBER OF INVENTION: GLYCOPROTEIN HORMONES
NUMBER OF SEQUENCES: 160
CORRESPONDENCE ADDRESS:

ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Avenue, Suite 200
STATE: Colorado

COUNTRY: USA

ZIP: 80111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MG
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/09472
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA: 08/488,402
APPLICATION NUMBER:
FILING DATE: 07 JUNE 1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/484,552
FILING DATE: 07 JUNE 1995
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX36-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 69 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
OTHER INFORMATION: All C's are 2'-NH2 modified
FEATURE:
OTHER INFORMATION: All U's are 2'-NH2 modified
PCT-US96-09472-16

Query Match 48.0%; Score 12; DB 5; Length 69;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 aggtgcacgcc 14
|||||
DB 25 AGGTGCATCGCC 14

RESULT 44
US-08-998-416-93/C
Sequence 93, Application US/08998416
GENERAL INFORMATION:
APPLICANT: Philippson, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Reibschung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYP11
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 93:
SEQUENCE CHARACTERISTICS:
LENGTH: 386 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1031UP
US-08-998-416-93

Query Match 48.0%; Score 12; DB 4; Length 386;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacgc 17
|||||
DB 292 TGCAATCGCCAGC 281

RESULT 45
US-08-743-637B-184/C
Sequence 184, Application US/08743637B
Patent No. 5994066
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: PICARD, Francois J.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESS:
ADDRESSEE: OUARLES & BRADY
STREET: 411 EAST WISCONSIN AVENUE
CITY: MILWAUKEE
STATE: WISCONSIN
COUNTRY: USA
ZIP: 53202-4497

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/743,637B
FILING DATE: 04-NOV-1996
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/526,840
FILING DATE: 11-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586, 90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000

```

; TELEFAX: (414)277-5591
; INFORMATION FOR SEQ ID NO: 184:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; US-08-743-637B-184

```

```

Query Match          48.0%; Score 12; DB 2; Length 438;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcagggtcaccg 12
   |||||||
Db 28 GCAggtcaccg 17

```

Search completed: December 26, 2001, 12:55:29
Job time: 6961 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:10:18 ; Search time 4619.78 Seconds
(without alignments)
58.151 Million cell updates/sec

Title: us-09-396-196f-6

Perfect score: 25
Sequence: 1 gccagtcgcatgccgcatcttgcgt 25

Scoring table: OLIGO-NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size: 9
Total number of hits satisfying chosen parameters: 509350

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database:

EST: *
1: em_estfun: *
2: em_esthum: *
3: em_estin: *
4: em_estom: *
5: em_estpl: *
6: em_estda: *
7: em_estrov: *
8: em_estov: *
9: em_hic: *
10: gb_est1: *
11: gb_est2: *
12: gb_hic: *
13: gb_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pln: *
18: em_gss_pro: *
19: em_gss_fod: *
20: em_gss_vrt: *
21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	17	68.0	498	10	A1533390 SD03891.5
C 2	16	64.0	376	11	BF933544 CM1-NT027
C 3	16	64.0	399	11	D24778 R1CR2545A R
C 4	16	64.0	429	10	AU030894 AU030894
C 5	16	64.0	510	10	AU031475 AU031475
C 6	16	64.0	548	13	P968R A1390674 Lelshman1
C 7	16	64.0	666	10	AU030895 AU030895
C 8	16	64.0	704	10	AU078750 AU078750
C 9	16	64.0	761	11	BF864751 BF864751
C 10	15	60.0	258	13	A2578088 A2578088
C 11	15	60.0	266	10	A1410504 A1410504
C 12	15	60.0	338	10	BE065518 BE065518

C 13	15	60.0	341	10	AV551712 AV551712
C 14	15	60.0	380	10	AA800788 AA800788
C 15	15	60.0	393	11	H60965 H60965
C 16	15	60.0	394	11	T86053 T86053
C 17	15	60.0	399	10	BE429060 BE429060
C 18	15	60.0	399	10	BE044431 BE044431
C 19	15	60.0	414	10	A1268648 A1268648
C 20	15	60.0	435	13	AA027122 AA027122
C 21	15	60.0	437	11	H69200 H69200
C 22	15	60.0	438	11	N92081 N92081
C 23	15	60.0	445	10	AV722315 AV722315
C 24	15	60.0	453	10	AU166733 AU166733
C 25	15	60.0	461	13	BE418794 BE418794
C 26	15	60.0	463	10	AA026650 AA026650
C 27	15	60.0	466	10	AA821080 AA821080
C 28	15	60.0	475	11	H60979 H60979
C 29	15	60.0	475	11	BG101817 BG101817
C 30	15	60.0	491	10	A1692369 A1692369
C 31	15	60.0	491	11	BF211097 BF211097
C 32	15	60.0	499	11	BF916376 BF916376
C 33	15	60.0	505	11	C98143 C98143
C 34	15	60.0	507	11	T84279 T84279
C 35	15	60.0	514	11	BS467085 BS467085
C 36	15	60.0	524	11	A1957283 A1957283
C 37	15	60.0	528	11	H74120 H74120
C 38	15	60.0	536	11	BG232022 BG232022
C 39	15	60.0	539	11	BF839030 BF839030
C 40	15	60.0	545	13	A2071920 A2071920
C 41	15	60.0	547	10	AM588600 AM588600
C 42	15	60.0	553	13	A2283519 A2283519
C 43	15	60.0	567	11	C98142 C98142
C 44	15	60.0	581	10	AA405993 AA405993
C 45	15	60.0	586	13	A2945591 A2945591
C 46	15	60.0	602	13	AO783596 AO783596
C 47	15	60.0	622	10	BE557935 BE557935
C 48	15	60.0	623	11	BG103212 BG103212
C 49	15	60.0	630	10	BE583417 BE583417
C 50	15	60.0	684	11	BF156891 BF156891
C 51	15	60.0	761	10	BE194330 BE194330
C 52	15	60.0	792	11	BG299490 BG299490
C 53	15	60.0	915	10	BF581354 BF581354
C 54	15	60.0	915	11	BG758882 BG758882
C 55	15	60.0	986	11	A2922752 A2922752
C 56	14	56.0	200	13	AM065325 AM065325
C 57	14	56.0	203	10	AA973902 AA973902
C 58	14	56.0	203	11	BG589383 BG589383
C 59	14	56.0	205	11	BG558891 BG558891
C 60	14	56.0	211	13	AO931496 AO931496
C 61	14	56.0	231	10	A1329089 A1329089
C 62	14	56.0	242	10	A1902416 A1902416
C 63	14	56.0	259	10	AA515233 AA515233
C 64	14	56.0	267	10	AA165335 AA165335
C 65	14	56.0	269	10	BA499261 BA499261
C 66	14	56.0	286	10	AA165335 AA165335
C 67	14	56.0	286	11	W49454 W49454
C 68	14	56.0	287	11	AM260563 AM260563
C 69	14	56.0	291	11	BE826582 BE826582
C 70	14	56.0	291	10	BB153277 BB153277
C 71	14	56.0	311	10	BB216989 BB216989
C 72	14	56.0	314	11	B1320345 B1320345
C 73	14	56.0	322	11	BG354207 BG354207
C 74	14	56.0	322	11	BF714260 BF714260
C 75	14	56.0	328	10	AU209015 AU209015
C 76	14	56.0	334	10	AA775899 AA775899
C 77	14	56.0	340	10	AW902605 AW902605
C 78	14	56.0	352	10	A1612263 A1612263
C 79	14	56.0	359	10	BE510828 BE510828
C 80	14	56.0	360	10	AV193175 AV193175
C 81	14	56.0	360	11	CA23256 CA23256
C 82	14	56.0	360	11	CA5784 CA5784
C 83	14	56.0	360	11	CA9684 CA9684
C 84	14	56.0	360	11	C50588 C50588
C 85	14	56.0	360	11	C50588 C50588

AV551712 AV551712
AA800788 EST190285
H60965 YR22607.r1
T86053 Y462405.r1
BE429060 MTBD013.H0
BE044431 h045d04.x
A1268648 q039f05.x
AA027122 C1T-HSP-2
H69200 YR59f10.r1
N92081 za23h03.r1
AV722315 AV722315
AU166733 AU166733
BH018794 L1975k.d
BE446670 WHE1139.F
AA026650 z199e08.r
AA821080 GM09622.5
H60979 YR22f01.r1
BG101817 RH122.22
A1692369 w063f02.x
BF211097 601812706
BF916376 IL3-UT011
C98143 C98143
T84279 Yd47b04.r1
BG467085 1A04C06.B
A1957283 u185e12.x
H74120 YS15e08.r1
BG232022 na135e06.
BF839030 PM2-HP035
A2071920 RPCI-23-3
AM588600 ra06g05.Y
A2283519 RPCI-23-1
C98142 C98142
AA405993 zu6f09.r
A2945591 2M0207H04
AO783596 HS-3121.A
BE557935 f115d06.Y
BG103212 RH122.19
BE583417 11-6H-MY
BF156891 f158b12.Y
BE194330 HVSME1002
BG299490 HVSME1002
A1282984 4A3A-P2G5
BF581354 602100740
BG758892 602713004
A2922752 S1C0T4A05
AM065325 614045H11
BG589383 EST197225
BG558891 RH122.57
AO931496 RPCI-23-2
A1329089 a9q12ne.r
A1902416 CM-BT006-
AA973902 CO47g09.S
AA515233 ng68b06.S
BA499261 BB499261
A165335 zq49c06.r
W49454 zEST01191.M
AM260563 um83c06.Y
BE826582 QV1-EN004
BB153277 BB153277
BB216989 BB216989
B1320345 sa121h02.
BG354207 947032F05
BF714260 mab35f04.
AU209015 AU209015
AA775899 z149f09.S
AM902605 QV3-NN102
A1612263 486078C11
BE510828 946055G06
AV193175 AV193175
CA23256 CA23256
CA5784 CA5784
CA9684 CA9684
C50588 C50588

```

c 86 14 56.0 362 10 AA872290 OH72302.S
c 87 14 56.0 369 11 M89214
c 88 14 56.0 375 10 AM902643
c 89 14 56.0 376 10 AV191473
c 90 14 56.0 384 10 AV195681
c 91 14 56.0 386 10 AM788306
c 92 14 56.0 387 10 BI295706
c 93 14 56.0 387 10 BI295706
c 94 14 56.0 387 10 BI295706
c 95 14 56.0 391 10 C00344-F
c 96 14 56.0 391 10 C00344-F
c 97 14 56.0 399 10 C00344-F
c 98 14 56.0 401 10 C00344-F
c 99 14 56.0 402 10 C00344-F
c 100 14 56.0 402 10 C00344-F

```

ALIGNMENTS

```

RESULT 1
LOCUS AI532390/c
DEFINITION SD03891.Sprime SD Drosophila melanogaster Schneider L2 cell culture
P072 Drosophila melanogaster CDNA clone SD03891 5 similar to
CG8473: Epan008473 'transporter' located on: X 13E18-13F1;
04/13/2001, mRNA sequence.
AI532390
AI532390.1 GI:4446525
EST.

```

```

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

```

```

Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 498)
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
Lewis, S., and Rubin, G.M.
BDGP/HHMI Drosophila EST Project
Unpublished (2001)
Other ESTs: SD03891.Sprime
Contact: Stapleton, M.
BDGP

```

```

FEATURES
source
Lawrence Berkeley National Lab
One Cyclotron Rd. Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST/estefruitfly.berkeley.edu
hit genomic A6003500: arm: X [15448919,15776364]
estimated-cyto: 13E14-14A4: 04/13/2001
Plate: SD 38 row: H column: 7
High quality sequence stop: 457
POLYA=No.

```

```

FEATURES
source
1. 498
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="SD03891"
/clone="SD03891"
/culture="SD Drosophila melanogaster Schneider L2 cell
culture port2"
/lab_host="DHS-alpha"
/notes="Vector: port2; Site_1: EcoRI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into port2.
Plasmid cDNA library."
BASE COUNT
158 a 106 c 124 g 110 t
ORIGIN

```

```

Query Match
Best Local Similarity 68.0%; Score 17; DB 10; Length 498;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
7 gcatccagcatctgc 23
|||||

```

```

Db 266 GCATGCCAGCATTTTCG 250

```

```

RESULT 2
LOCUS BF933544/c
DEFINITION CMT-NT0270-211200-661.903 NT0270 Homo sapiens CDNA, mRNA sequence.
BF933544
ACCESSION BF933544.1 GI:12350868
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 376)
Dias Neto, E., Garcia Correa, R., Verjovsky-Almeida, S., Brites, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Rue Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the RAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM1&lt2=CM1-NT0270-
211200-661.903&lt3=2000-12-21&lt4=1)
Seq primer: puc 18 forward
High quality sequence start: 10
High quality sequence stop: 376.

```

```

FEATURES
source
1. 376
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NT0270"
/dev_stage="Adult"
/notes="Organ: nervous tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

```

BASE COUNT
77 a 104 c 99 g 96 t
ORIGIN

```

```

Query Match
Best Local Similarity 64.0%; Score 16; DB 11; Length 376;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
4 gctgcatccagcat 19
|||||
Db 329 GCTGCATCCAGCAT 314

```

```

RESULT 3
LOCUS D24778
DEFINITION R1CR2545A Rice root Oryza sativa CDNA clone R2545_1A, mRNA
sequence.
ACCESSION D24778
VERSION D24778.1 GI:428626

```

```

KEYWORDS      EST
SOURCE         Oryza sativa.
ORGANISM       Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
                Euhartioideae; Oryzaceae; Oryza.
REFERENCE      1 (bases 1 to 399)
                Minobe,Y. and Sasaki,T.
                Rice cDNA from root
                Unpublished (1995)
                Contact: Takuji Sasaki
                National Institute of Agrobiological Resources
                Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
                305-8602, Japan
                Tel: 81-298-38-7441
                Fax: 81-298-38-7468
                Email: tsasaki@abr.affrc.go.jp, URL:http://rpg.dna.affrc.go.jp/
                PROJECT="RGP"
FEATURES       Location/Qualifiers
                source          1..399
                /organism="Oryza sativa"
                /strain="Nipponbare, sub-species Japonica"
                /db_xref="taxon:4530"
                /clone.lib="R2545_1A"
                /clone.lib="Rice root"
                /note="Prepared from seedling root."
BASE COUNT    114 a      70 c      102 g      112 t      1 others
ORIGIN
Query Match   64.0%; Score 16; DB 11; Length 399;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY            4 ggtcctcgcgcagcat 19
                |||
Db            251 GGTCATCCGCGCAGCAT 266

RESULT 4
LOCUS      AU030894      429 bp      mRNA      EST      29-OCT-1998
DEFINITION AU030894 Rice cDNA from immature leaf including apical meristem
ACCESSION  AU030894
VERSION     AU030894.1 GI:376784
KEYWORDS    EST.
SOURCE      Oryza sativa.
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
                Euhartioideae; Oryzaceae; Oryza.
                1 (bases 1 to 429)
                Sasaki,T. and Yamamoto,K.
                Rice cDNA from immature leaf including apical meristem
                Unpublished (1997)
                Contact: Takuji Sasaki
                National Institute of Agrobiological Resources
                Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
                305-8602, Japan
                Tel: 81-298-38-7441
                Fax: 81-298-38-7468
                Email: tsasaki@abr.affrc.go.jp, URL:http://rpg.dna.affrc.go.jp/
                PROJECT="RGP"
FEATURES     Location/Qualifiers
                source          1..429
                /organism="Oryza sativa"
                /strain="Nipponbare"
                /db_xref="taxon:4530"
                /clone.lib="E60405_1A"
                /clone.lib="Rice cDNA from immature leaf including apical
                meristem"
                /dev_stage="immature"

```

```

/note="Organ: leaf; immature leaf including apical
meristem (under long day condition)"
BASE COUNT    106 a      122 c      101 g      99 t      1 others
ORIGIN
Query Match   64.0%; Score 16; DB 10; Length 429;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY            8 catgccagcatctcg 23
                |||
Db            302 CATGCCAGCATTTTCG 287

RESULT 5
LOCUS      AU031475      510 bp      mRNA      EST      30-OCT-1998
DEFINITION AU031475 Rice cDNA from immature leaf including apical meristem
ACCESSION  AU031475
VERSION     AU031475.1 GI:3767365
KEYWORDS    EST.
SOURCE      Oryza sativa.
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
                Euhartioideae; Oryzaceae; Oryza.
                1 (bases 1 to 510)
                Sasaki,T. and Yamamoto,K.
                Rice cDNA from immature leaf including apical meristem
                Unpublished (1997)
                Contact: Takuji Sasaki
                National Institute of Agrobiological Resources
                Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
                305-8602, Japan
                Tel: 81-298-38-7441
                Fax: 81-298-38-7468
                Email: tsasaki@abr.affrc.go.jp, URL:http://rpg.dna.affrc.go.jp/
                PROJECT="RGP"
FEATURES     Location/Qualifiers
                source          1..510
                /organism="Oryza sativa"
                /strain="Nipponbare"
                /db_xref="taxon:4530"
                /clone.lib="E61659_22"
                /clone.lib="Rice cDNA from immature leaf including apical
                meristem"
                /dev_stage="immature"
                /note="Organ: leaf; immature leaf including apical
                meristem (under long day condition)"
BASE COUNT    176 a      98 c      99 g      137 t
ORIGIN
Query Match   64.0%; Score 16; DB 10; Length 510;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY            8 catgccagcatctcg 23
                |||
Db            136 CATGCCAGCATTTTCG 121

RESULT 6
LOCUS      P968R/C      548 bp      DNA      GSS      25-JUL-2000
DEFINITION P968R Leishmania major Friedlin PAC P968 right end-sequence, genomic
                survey sequence.
ACCESSION  AL390674
VERSION     AL390674.1 GI:9501650
KEYWORDS    GSS.
SOURCE      Leishmania major.

```

```

ORGANISM      Leishmania major
               Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
REFERENCE      1 (bases 1 to 548)
AUTHORS        Ivens,A.C., Lewis,S.M., Bagherzadeh,A., Zhang,L., Chan,H.M. and
               Smith,D.F.
TITLE          A physical map of the Leishmania major Friedlin genome
JOURNAL        Genome Res. 8 (2), 135-145 (1998)
MEDLINE        98146435
REFERENCE      2 (bases 1 to 548)
AUTHORS        Taylor,R.G., Huckle,E.F.J., Ivens,A.C., Rajandream,M.A. and
               Barrell,B.G.
TITLE          Direct Submision
JOURNAL        Submitted (24-JUL-2000) Leishmania major Friedlin genome sequencing
               project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
               Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
               allcats@sanger.ac.uk
               see http://www.ebi.ac.uk/parasites/leish.html
               Details of Leishmania sequencing at the Sanger Centre are available
               at http://www.sanger.ac.uk/projects/L_majior/
               The primer sequence can be obtained from allcat@sanger.ac.uk.
FEATURES       Location/Qualifiers
               1..548
               /organism="Leishmania major"
               /strain="Friedlin"
               /db_xref="taxon:5664"
               /clone="PAC p968"
BASE COUNT     86 a 167 c 163 g 132 t
ORIGIN
Query Match    64.0%; Score 16; DB 13; Length 548;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 ggcgtcgccagcatt 20
Db 541 gTGCATCGCCAGCATT 526

RESULT 7
LOCUS      AU030895/c 666 bp mRNA EST 29-OCT-1998
DEFINITION AU030895 Rice cDNA from immature leaf including apical meristem
ACCESSION  AU030895
VERSION     AU030895.1 GI:3766785
KEYWORDS    EST.
SOURCE      Oryza sativa.
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
               Eriharitoidae; Oryzaceae; Oryza.
               1 (bases 1 to 666)
               Sasaki,T. and Yamamoto,K.
               Rice cDNA from immature leaf including apical meristem
REFERENCE    Unpublished (1997)
AUTHORS      Contact: Takuji Sasaki
TITLE        National Institute of Agrobiological Resources
JOURNAL      Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
               305-8602, Japan
               Tel: 81-298-38-7441
               Fax: 81-298-38-7468
               Email: tsasaki@abr.affrc.go.jp, URL:http://rpg.dna.affrc.go.jp/
               PROJECT="RGP".
FEATURES     Location/Qualifiers
               1..666
               /organism="Oryza sativa"
               /strain="Nipponbare"
               /db_xref="taxon:4530"
               /clone="E60405_22"
               /clone_lib="Rice cDNA from immature leaf including apical
               meristem"

```

```

/dev_stage="Immature"
/Note="Organ: leaf; Immature leaf including apical
meristem (under long day condition)"
BASE COUNT     195 a 149 c 131 g 181 t 10 others
ORIGIN
Query Match    64.0%; Score 16; DB 10; Length 666;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 catcgccagcattcg 23
Db 293 CATCGCCAGCATTTCG 278

RESULT 8
LOCUS      AU078750/c 704 bp mRNA EST 18-OCT-1999
DEFINITION AU078750 Rice panicle at flowering stage Oryza sativa cDNA clone
ACCESSION  AU078750
VERSION     AU078750.1 GI:6062509
KEYWORDS    EST.
SOURCE      Oryza sativa.
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
               Eriharitoidae; Oryzaceae; Oryza.
               1 (bases 1 to 704)
               Sasaki,T. and Yamamoto,K.
               Rice cDNA from panicle at flowering stage
REFERENCE    Unpublished (1996)
AUTHORS      Contact: Takuji Sasaki
TITLE        National Institute of Agrobiological Resources
JOURNAL      Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
               305-8602, Japan
               Tel: 81-298-38-7441
               Fax: 81-298-38-7468
               Email: tsasaki@abr.affrc.go.jp, URL:http://rpg.dna.affrc.go.jp/
               PROJECT="RGP".
FEATURES     Location/Qualifiers
               1..704
               /organism="Oryza sativa"
               /strain="Nipponbare"
               /db_xref="taxon:4530"
               /clone="E0723_42"
               /clone_lib="Rice panicle at flowering stage"
               /dev_stage="flowering stage"
               /note="Organ: panicle; Rice cDNA from panicle at flowering
               stage"
BASE COUNT     203 a 167 c 142 g 191 t 1 others
ORIGIN
Query Match    64.0%; Score 16; DB 10; Length 704;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 catcgccagcattcg 23
Db 346 CATCGCCAGCATTTCG 331

RESULT 9
LOCUS      BF864751 761 bp mRNA EST 19-JAN-2001
DEFINITION BF864751 C. reinhardtii CC-1690, Stress condition I, normalized
               Lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION  BF864751
VERSION     BF864751.1 GI:12254895
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.

```


ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonas.

REFERENCE 1 (bases 1 to 761)
Grossman, A., Davies, J., Federspiel, N., Harris, E., Hauser, C.,
Lefebvre, P., McDermott, J.P., Shriver, J., Silljow, C. and Steen, D.
Analyses of the Chlamydomonas reinhardtii genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants: project phase 3
Unpublished (2000)

JOURNAL Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu

FEATURES
source
1..761
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, stress condition 1,
normalized, lambda Zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr
, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). POLYA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into lambda Zap II
(Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
Zap clones by superinfection with EXAssist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 153 a 179 c 308 g 117 t 4 others

ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 761;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 aggtcgcgcgcagca 18
|||||

Db 466 AGGTGATCGCCAGCA 451

RESULT 10
A2578088 258 bp DNA GSS 08-DEC-2000
LOCUS 18f02 Shot-gun genomic library of Rhizobium strain ANU265 Rhizobium
DEFINITION sp. NGR234 genomic clone 18f02, DNA sequence.
A2578088
A2578088.1 GI:11605015
GSS.
Rhizobium sp. NGR234.
Rhizobium sp. NGR234
Bacteria; Proteobacteria; alpha subdivision: Rhizobiaceae group;
Rhizobiaceae; Rhizobium.
1 (bases 1 to 258)
Viprey, V., Rosenthal, A., Broughton, W.J. and Perret, X.
Genetic snapshots of the Rhizobium species NGR234 genome
Genomebiology.com 1 (6), 0014.1-0014.7 (2000)
Contact: Virginie Viprey
Laboratoire de Biologie Molculaire des Plantes Supérieures
University of Geneva
1 Chemin de l'Imperatrice, Chambesey/Geneva 1292, Switzerland
Tel: +44(0)1603450000
Fax: +44(0)1603450045
Email: virginie.viprey@bbsrc.ac.uk

Class: shotgun.
FEATURES
source
Location/Qualifiers
1..258
/organism="Rhizobium sp. NGR234"
/strain="ANU265"
/db_xref="taxon:394"
/clone_lib="Shot-gun genomic library of Rhizobium strain
ANU265"
/note="Vector: M13; derivative strain of NGR234 cured of
pNGR234a"

BASE COUNT 51 a 82 c 79 g 46 t

ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 258;
Best Local Similarity 100.0%; Pred. No. 11e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tcgcgcgcattcga 24
|||||

Db 175 TCGCCAGCATTTTCCA 161

RESULT 11
A1410504 266 bp mRNA EST 09-FEB-1999
LOCUS EST238797 Normalized rat heart, Bento Soares Rattus sp. cDNA clone
DEFINITION RHECX29 3' end, mRNA sequence.
A1410504
A1410504.1 GI:4254008
EST.
Rattus sp.
Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 266)
Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
Kerlavage, A.R. and Adams, M.D.
Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
Gene Index
Unpublished (1998)
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
1..266
/organism="Rattus sp."
/db_xref="taxon:10118"
/clone_lib="RHECX29"
/clone_lib="Normalized rat heart, Bento Soares"
/note="Organ: heart; Vector: pRTT3pac; Site_1: EcoRI;
Site_2: NotI"

BASE COUNT 56 a 61 c 64 g 85 t

ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 266;
Best Local Similarity 100.0%; Pred. No. 11e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ggtcgcgcgcagca 18
|||||

Db 90 GGTGATCGCCAGCA 104

RESULT 12
BE06518/c

```

LOCUS       BE065518             338 bp    mRNA                      EST          09-JUN-2000
DEFINITION  RC3-BT0316-170200-014-b06 BT0316 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE065518
VERSION     BE065518.1   GI:8410168
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE   1 (bases 1 to 338)
AUTHORS     Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brites, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
TITLE       Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE    20202663
COMMENT     Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/ICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=6&t2=RC3-BT0316-170
200-014-b06&t3=2000-02-17&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 338.
FEATURES
    source
        1..338
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_id="BT0316"
            /dev_stage="Adult"
            /note="Organ: breast; Vector: puc18; Site_1: Sma1; Site_2:
            from ORESTES PCR (U.S. Letters Patent application No. 196
            ,716 - Ludwig Institute for Cancer Research) profiles
            into the pUC 18 vector. Reverse transcription of tissue
            mRNA and cDNA amplification were performed under low
            stringency conditions."
BASE COUNT
    76 a      100 c      66 g      96 t

Query Match
Best Local Similarity 100.0%; Score 15; DB 10; Length 338;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8  catgccagcatctc 22
        |||
Db      97  CATGCCAGCATTTTC 83

RESULT 13
LOCUS     AV551712             341 bp    mRNA                      EST          06-SEP-2000
DEFINITION  AV551712 Arabidopsis thaliana roots Columbia
ACCESSION   AV551712
VERSION     AV551712.1   GI:8723125
KEYWORDS    EST.
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

FEATURES
    source
        1..380
            /organism="Arabidopsis thaliana"
            /Strain="Columbia"
            /db_xref="taxon:3702"
            /clone_id="R213e05R"
            /clone_lib="Arabidopsis thaliana roots Columbia"
            /tissue_type="roots"
            /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
            XhoI"
BASE COUNT
    73 a      85 c      53 g      130 t

Query Match
Best Local Similarity 100.0%; Score 15; DB 10; Length 341;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      10  tcgcagcatctcga 24
        |||
Db      322  TCCGACGACATTTGCA 336

RESULT 14
LOCUS     AA800788             380 bp    mRNA                      EST          30-APR-1998
DEFINITION  EST190285 Normalized rat lung, Bento Soares Rattus sp. cDNA clone
ACCESSION   AA800788
VERSION     AA800788.1   GI:2863743
KEYWORDS    EST.
SOURCE      Rattus sp.
ORGANISM    Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE   1 (bases 1 to 380)
AUTHORS     Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
Kerlavage, A.R. and Adams, M.D.
TITLE       Rat Genome Project: Generation of a Rat EST (RESt) Catalog & Rat
Gene Index
JOURNAL     Unpublished (1998)
COMMENT     Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
FEATURES
    source
        1..380
            /organism="Rattus sp."
            /db_xref="ATCC (inhost):2008206"
            /db_xref="taxon:10118"
            /clone_id="R213e05R"
            /clone_lib="R213e05R"
            /note="Organ: lung; Vector: pT73pac; Site_1: EcoRI;
            Site_2: NotI"
BASE COUNT
    71 a      100 c      106 g      103 t

```


[illegible]

```

201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@cgic.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
    source
        1..445
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="HBAMG01"
            /clone_1lb="HTB"
            /tissue_type="Hypothalamus"
            /dev_stage="Adult"
            /lab_host="SOLR"
            /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
            XhoI"
BASE COUNT      102 a      123 c      104 g      116 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 15; DB 10; Length 445;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY      8 catgccagcatttc 22
        |||||||
Db      296 CATGCCAGCATTTG 282

RESULT 24
LOCUS   AUI66733 453 bp mRNA EST 26-APR-2001
DEFINITION AUI66733 Rice callus (2001) Oryza sativa CDNA clone C50066, mRNA
sequence.
ACCESSION AUI66733
VERSION   AUI66733.1 GI:12405132
KEYWORDS  EST.
SOURCE    Oryza sativa.
ORGANISM  Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 453)
Sasaki, T. and Yamamoto, K.
Rice cDNA from callus (2001)
Unpublished (2001)
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tssasaki@dr.afrc.go.jp, URL: http://rtp.dna.afrc.go.jp/
PROJECT "RGP".
C50066_982.
FEATURES
    source
        1..453
            /organism="Oryza sativa"
            /cultivar="Nipponbare"
            /db_xref="taxon:4530"
            /clone="C50066"
            /clone_1lb="Rice callus (2001)"
            /tissue_type="callus"
BASE COUNT      164 a      163 c      140 g      66 t
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 15; DB 10; Length 453;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
        6 tgcattgccagcatt 20
        |||||||

```

Db 349 TGCGATGCCGCGCAT 335

RESULT 25
LOCUS BH018794/c
DEFINITION L1975K.d.Hy973.1 Leishmania major Friedlin Cosmid Genomic Library
Leishmania major genomic clone L1975K, DNA sequence.

ACCESSION
VERSION BH018794.1 GI:14197500
KEYWORDS
SOURCE
ORGANISM Leishmania major.
Eukaryota; Eukaryota; Kinetoplastida; Trypanosomatidae; Leishmania.

REFERENCE
AUTHORS 1 (bases 1 to 461)
Myler,P.J., Vogt,C., Gauthra,J., Kieckhefer,M., Marty,A., Mack,J., Munden,H., Nguyen,D., Robertson,L., Sisk,E., Fazelinia,G., Aggarwal,G., Nelson,S., Seyler,A., Worthey,E. and Stuart,K.
Leishmania major Friedlin Cosmid End Sequences
Unpublished (2000)
Other GSS: L1975K.d.Hy973a.1
Contact: Myler PJ
Seattle Biomedical Research Institute
4 Nickerson Street, Seattle, WA 98109-1651, USA
Tel: 206 284-8846
Fax: 206 284-0313
Email: mylerpj@sbri.org
Seq primer: Hy973
Class: cosmid ends.

FEATURES
source
1..461
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/clone="L1975K"
/clone_1lb="Leishmania major Friedlin Cosmid Genomic Library"
/lab_host="E. coli ED8767"
/note="Vector: cLHYG; Site_1: BamHI; Genomic DNA from Leishmania major Friedlin was partially digested with Sma3AT, size selected, and ligated with BamHI-digested cLHYG cosmid vector DNA. 9216 clones were picked and arrayed. Library construction is described in Irens et al., Genomics Research, 8:135-145 (1998). The cLHYG vector (ACC. No. CVU59231) is described in Ryan et al., Gene, 131:145-150 (1993)"

BASE COUNT 100 a 130 c 129 g 101 t 1 others

ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 461;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 caggtgcacgcagcag 16
|||||
Db 279 CAGGTGCATGCCGCG 265

RESULT 26
LOCUS BE446670/c
DEFINITION BE446670 463 bp mRNA EST 25-JUL-2000
WHE1139_F07_L13S Wheat etiolated seedling root normalized cDNA library Triticum aestivum cDNA clone WHE1139_F07_L13, mRNA sequence.

ACCESSION
VERSION BE446670.1 GI:9446232
KEYWORDS
SOURCE
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae

REFERENCE
AUTHORS 1 (bases 1 to 463)
Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Nguyen,H.T., Rausch,C.J., Seaton,C.L., Tong,J.C. and Zhang,D.
The structure and function of the expressed portion of the wheat genomes - Normalized root cDNA library
Unpublished (2000)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@w.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: StrataGene SK primer.

FEATURES
source
1..463
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE1139_F07_L13"
/clone_1lb="Wheat etiolated seedling root normalized cDNA library"
/tissue_type="Root"
/dev_stage="Five day old etiolated seedling"
/lab_host="E. coli DH10B"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid phiscript SK; Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized, germinated and grown aseptically in the dark at room temperature in covered crystallization dishes. Roots were harvested. The tissue, total RNA, and poly(A) RNA were prepared, a cDNA library was made in the T7 Close lab (Choi, Close, Fenton) at the University of California, Riverside. The cDNA clones were in vivo excised to give phiscript phagemids before normalization was carried out. The mass excision of phagemid library and normalization were done in HT Nguyen lab by D. Zhang at Texas Tech University. Normalization protocol used was that of Soares. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT 119 a 98 c 132 g 114 t

ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 463;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcagtgcatgcgca 15
|||||
Db 111 GCAGTGCGATGCCGCA 97

RESULT 27
LOCUS AA026650/c
DEFINITION 2J99e08.r1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone IMAGE:469190 5', mRNA sequence.

ACCESSION
VERSION AA026650
KEYWORDS
SOURCE
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 466)
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chapelil,B., Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins

with a modified polylinker. Site-1: Pac I; Site-2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGCAGAGATTAAATTAAAGATCTTTTCTTTT 3'], (double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pTV73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 109 a 123 c 95 g 143 t 5 others
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 475;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgccagcatc 22
|||||
Db 119 CATGCCAGCATTTTC 105

RESULT 30
BG101817 475 bp mRNA EST 30-JAN-2001
LOCUS BG101817
DEFINITION RH122.22.G09.g1.A003 Rhizome2 (RH122) Sorghum prolinguim cDNA, mRNA
sequence.
ACCESSION BG101817
VERSION BG101817.1 GI:12616762
KEYWORDS EST.
SOURCE Sorghum prolinguim.
ORGANISM Sorghum prolinguim.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 475)
Cordonnier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt,
L.H.

REFERENCE 1
AUTHORS Cordonnier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt,
L.H.

TITLE An EST database from Sorghum: Sorghum prolinguim rhizomes
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Seq primer: PolyTMix
High quality sequence start: 11
High quality sequence stop: 408
POLYA-No.

FEATURES
source Location/Qualifiers
1..475
/organism="Sorghum prolinguim"
/db_xref="taxon:132711"
/clone_lib="Rhizome2 (RH122)"
/note="Organ: Rhizomes; Vector: pBluescript II from lambda
zap II; Site 1: XhoI; Site 2: EcoRI; The library was made
from poly(A RNA in the cloning vector lambda zap II.
Clones that were sequenced were prepared by mass excision."

BASE COUNT 128 a 107 c 113 g 127 t
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 475;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcagtgacgcgca 15
|||||
Db 240 GCAGTGACGCCCA 254

RESULT 31

AI692369
LOCUS AI692369 491 bp mRNA EST 17-DEC-1999
DEFINITION wd63f02.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2336283 3',
mRNA sequence.
ACCESSION AI692369
VERSION AI692369.1 GI:4969709
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 491)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrp/image/image.html
Insert length: 589 Std Error: 0.00
Seq primer: -40up from Gibco
High quality sequence stop: 471.

FEATURES
source Location/Qualifiers
1..491
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:2336283"
/clone_lib="NCI CGAP_Lu24"
/issue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: Lung; Vector: pTV73-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI CGAP_Lu25 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (clones
141920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 151 a 94 c 129 g 116 t 1 others
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 491;
Best Local Similarity 100.0%; Pred. NO. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgccagcatc 22
|||||
Db 344 CATGCCAGCATTTTC 358

RESULT 32
BF211097 491 bp mRNA EST 06-NOV-2000
LOCUS BF211097
DEFINITION 601812706f1 NIH_MGC_54 Homo sapiens cDNA clone IMAGE:4047120 5',
mRNA sequence.
ACCESSION BF211097
VERSION BF211097.1 GI:11104683
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 491)
NIH-MGC http://mgs.nci.nih.gov/.

REFERENCE
AUTHORS

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLM875 row: 1 column: 01
High quality sequence start: 56
High quality sequence stop: 491.
Location/Qualifiers
1. 491
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4047120"
/clone_lib="NIH MGC 54"
/tissue_type="From chronic myelogenous leukemia"
/note="Organ: Bone marrow; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccgctcgcc); Site_2: SfiI (ggccatagcc 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-ATCTAGAGCCGCGCCGCGCCATG-3' and 3' adaptor (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."
BASE COUNT 123 a 128 c 119 g 121 t
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 491;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catccgcagcatttc 22
|||||
Db 305 CATCCGCGCATTTTC 291

RESULT 33
BF916376 499 bp mRNA EST 18-JAN-2001
LOCUS IT3-UT0115-111200-377-F05 UT0115 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF916376
ACCESSION BF916376
VERSION BF916376.1 GI:12307834
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 499)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovsky-Almolda,S., Briones,M.R., Nagai,M.A., da Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=113&cl=113-UT0115-111200-377-F05&cl3=2000-12-11&cl4=1>)
Seq primer: puc 18 forward
High quality sequence stop: 496.
Location/Qualifiers
1. 499
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="UT0115"
/dev_stage="Adult"
/note="Organ: uterus-tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196/716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 121 a 121 c 114 g 143 t
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 499;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaagtcagtcgcca 15
|||||
Db 216 GCAAGTCATCGCCA 230

RESULT 34
C98143/C 505 bp mRNA EST 29-APR-1999
LOCUS C98143 Rice callus Oryza sativa cDNA clone C0777_7A, mRNA sequence.
DEFINITION C98143 D28202
ACCESSION C98143
VERSION C98143.1 GI:3760889
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 505)
AUTHORS Yamamoto,K. and Sasaki,T.
TITLE Rice cDNA from callus (1998)
JOURNAL Unpublished (1998)
COMMENT On Oct. 19, 1998 this sequence version replaced gi:454479.
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@ab.affrc.go.jp, URL: <http://rgp.dna.affrc.go.jp/D28202>. Submitted (01-JAN-1994)
PROJECT "RGP"
POLY-A=NO.
Location/Qualifiers
1. 505
/organism="Oryza sativa"
/strain="Cultivar Nipponbare, sub-species Japonica"
/db_xref="taxon:4530"
/clone_image="C0777_7A"
/clone_lib="Rice callus"
/note="Vector: Bluescript II SK-; Site_1: SalI; Site_2: NotI; cDNA prepared from rice callus mRNAs by using oligo(dT) as a primer and ligating to the SalI-NotI site of phuescript II SK+ phagemid."

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euteria; Rodentia; Sclurognathi; Muridae; Murinae; Mus 1 (bases 1 to 524)

Waterston, R. and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through INL; contact the
IMAGe Consortium (info@image.inl.gov) for further information.
MG1:1002066
Seq primer: custom primer used
High quality sequence stop: 359.
Location/Qualifiers
1..524

```

/dev_stage="adult"
/lab_host="DH10b"
/note="Organ: kidney; Vector: pME18S-F13; Site_1: DraII
(CACCTGCTG); Site_2: DraII (CACCACTGTG); 1st strand cDNA
was primed with an oligo(dT) primer
(ATATGCGCCCTTTTATTTTTTTTTTTT); double-stranded cDNA was
ligated to a DraII adaptor (TCTGGCGCTACCTGG), digested
and cloned into distinct DraII sites of the pME18S-F13
vector (5' site CACCTGCTG, 3' site CACCACTGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTCCTGCTCTTAAACCTGCG and 3' end
primer CACCTGCGACCTGAGGACA."
151 a 102 c 147 g 124 t

```

Query	6	TGATCGCAGCATT	20
Best Local Similarity	60.0%;	Score 15;	DB 10; length 524;
Matches	15;	Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Db	352	TGATCGCAGCATT	338

RESULT	38
H74120/c	
LOCUS	H74120
DEFINITION	528 bp mRNA
ACCESSION	U91560.f1 Soares fetal liver spleen INFLS Homo sapiens EST
VERSION	IMAE:214886.5', mRNA sequence.
KEYWORDS	H74120.1 GI:1047332
EST	
SOURCE	human.
ORGANISM	Homo sapiens
	31-CCN-1995 cDNA clone

boarivola; Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi; Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo, 1 (bases 1 to 528)

Hillier, L., Lemon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Marids, E., Moore,

B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Terry-Meg, J., Trevisan, E., Underwood, K., Wohlmann, P., Waterson, R., Wilson, R. and Marra, M. Generation and analysis of 280,000 human expressed sequence tags Genome Res. 6 (9), 807-828 (1996)

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 1036
 High quality sequence stops: 311
 Source: IMAGE Consortium, LNM.
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1036 Std Error: 0.00
 Seq primer: M13RPI
 High quality sequence stop: 311.
 Location/Qualifiers
 1..528

121 a 124 c 110 g 166 t 7 others

Query Match	60.0%	Score 15;	DB 11;	Length 526;
Best Local Similarity	100.0%;	Pred. No. 1.1e+02;		
Matches 15; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	8	cacgcccaagcatctc	22
Db	107	CATCGCCAGCATTTTC	93
RESULT	39		
LOCUS	BG232022	536 bp	mRNA
DEFINITION	na435ae06.x1 Soares_NPBMHC Homo sapiens CDNA clone IMAGEF.4143059.3'		EST
			09-FEB-2001

LOCUS	BG232022	536 bp	mRNA	EST	09-FEB-2001
DEFINITION	na155e06.x1 Soares_PBMK Homo sapiens cDNA clone IMAGE:4143059 3'				
ACCESSION	/CYTOSPLASMIC DOMAIN ; mRNA sequence.				
VERSION	BG232022				
KEYWORDS	BG232022.1	GI:12727177			
SOURCE	EST.				
ORGANISM	human.				
	Homo sapiens				

Memorial Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 contact: Robert Strausberg, Ph.D.
 Email: cgapsr@mail.nih.gov
 Issue Procurement: M. Bento Soares, Ph.D.
 DNA Library Preparation: M. Bento Soares, Ph.D.
 and M. Telling

EcORI, Site-2: EcORI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcORI and EcORI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcORI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). "

BASE COUNT 137 a 153 c 121 g 134 t

ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcatgcga 15
|||||

Db 505 GCAGGTGATCGCA 491

RESULT 42
AM588900/c 547 bp mRNA EST 10-MAY-2001
LOCUS ra06905.y2 Bird-Rao Meloidogyne incognita J2 Meloidogyne incognita
DEFINITION cDNA 5' similar to TR:055013 055013 BET3. ; contains Alu repetitive element; mRNA sequence.
ACCESSION AM588900.1 GI:7275932
VERSION EST
KEYWORDS southern root-knot nematode.
SOURCE Meloidogyne incognita
ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina; Tylenchoidea; Heteroderidae; Meloidogyminae; Meloidogyne.

REFERENCE 1 (bases 1 to 547)
AUTHORS McCarter,J., Clifton,S., Chiappelli,B., Pape,D., Martin,J., Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarisvill,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schuk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterson,R. and Wilson,R.
TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: McCarter, JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
The library was constructed by Uma Rao and David Bird (david_bird@wustl.edu) at North Carolina State University. DNA sequencing by: Washington University Genome Sequencing Center St. Louis.
Seq primer: T3 ET from Amersham
High quality sequence stop: 519.

FEATURES

source location/Qualifiers
1. 547
/organism="Meloidogyne incognita"
/db_xref="taxon:6306"
/clone_lib="Bird-Rao Meloidogyne incognita J2"
/dev_stage="enriched for 2nd stage juveniles"
/lab_host="XLR1"

/note="Vector: ZAP express - pBACW (Stratagene); Site-1: EcORI; Site-2: XhoI; Oligo (dT) primed library. cDNA was constructed and cloned unidirectionally into the vector within the 5' EcORI and 3' XhoI sites. This library was constructed by Dr. Uma Rao and Dr. David Bird at North Carolina State University."

BASE COUNT 179 a 76 c 106 g 185 t 1 others

ORIGIN

Query Match

60.0%; Score 15; DB 10; Length 547;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgcgcagcattc 22
|||||

Db 296 CATGCAGCATTC 282

RESULT 43
A2283519 553 bp DNA GSS 27-JUL-2000
LOCUS RPCI-23-125E16.TU RPCI-23 Mus musculus genomic clone RPCI-23-125E16
DEFINITION , DNA sequence.
ACCESSION A2283519
VERSION A2283519.1 GI:9525226
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 553)
AUTHORS Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S., Akhret,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-23-125E16.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Reseach Genetics (Info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac/ends/mouse/bac_end_intro.html
Plate: 125 Row: E Column: 16
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1. 553
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="RPCI-23-125E16"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"

FEATURES

source location/Qualifiers
1. 553
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="RPCI-23-125E16"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site-1: EcORI; Site-2: EcORI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcORI and EcORI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcORI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). "

BASE COUNT 135 a 156 c 126 g 136 t

ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcatgcga 15
|||||

Db 525 GCAGGTGATCGCA 511

RESULT 44

C98142/c
LOCUS C98142 567 bp mRNA EST 19-OCT-1998
DEFINITION C98142 Rice callus Oryza sativa cDNA clone C0777_102, mRNA
sequence.
C98142
ACCESSION C98142
VERSION C98142.1 GI:3760888
KEYWORDS EST.
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 567)
AUTHORS Sasaki,T. and Minobe,Y.
TITLE Rice cDNA from callus
JOURNAL Unpublished (1994)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@agr.affrc.go.jp, URL: http://rtp.dna.affrc.go.jp/
PROJECT = RGP
FEATURES
source Location/Qualifiers
1..567
/organism="Oryza sativa"
/strain="cultivar Nipponbare, sub-species japonica"
/db_xref="taxon:4530"
/clone="C0777_102"
/clone_lib="Rice callus"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid. 1 others
BASE COUNT 190 a 105 c 196 g 75 t
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 567;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tgcacgcacgacatc 20
|||||
DB 472 TGCATCGCAGACATT 458

RESULT 45
AA405993/c 581 bp mRNA EST 09-NOV-1997
LOCUS AA405993
DEFINITION zu66f09.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:742985
5' similar to contains MER29.b3 MER29 repetitive element ;, mRNA
sequence.
AA405993
ACCESSION AA405993
VERSION AA405993.1 GI:2063976
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 581)
AUTHORS Hillier,L., Allen,K., Bowles,L., Dubuque,T., Geisels,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theisling,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

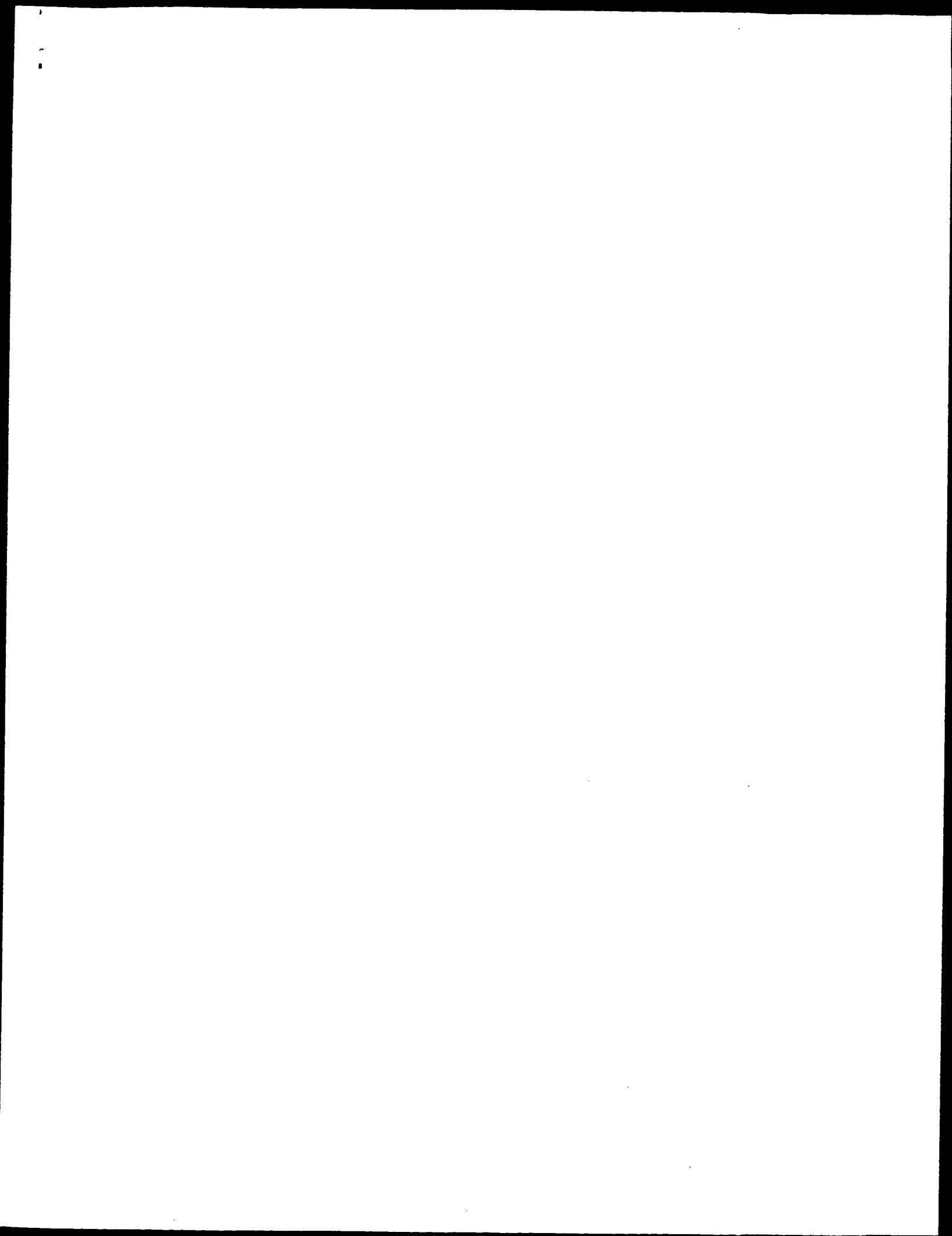
Email: est@watson.wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 997 Std Error: 0.00
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 498.
Location/Qualifiers
1..581
/organism="Homo sapiens"
/db_xref="GDB:5930030"
/db_xref="taxon:9606"
/clone="IMAGE:742985"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5'
TGTTCACCAATCTGAAGCGGAGCGCCGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Col5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 153 a 171 c 131 g 126 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 581;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgcacgacatc 22
|||||
DB 81 CATCGCAGACATTTC 67

Search completed: December 26, 2001, 14:10:24
Job time: 7071 sec




```

71 14 56.0 363 1 AB026474
72 14 56.0 372 1 AB026469
73 14 56.0 384 1 AB026470
74 14 56.0 390 1 AB026475
75 14 56.0 399 1 AB026458
76 14 56.0 426 1 AB026457
77 14 56.0 432 1 AB026455
78 14 56.0 435 1 AB026460
79 14 56.0 450 1 AB026456
80 14 56.0 614 6 AX154673
81 14 56.0 629 11 AX154673
82 14 56.0 1101 1 BMS2037
83 14 56.0 1135 8 AF291659
84 14 56.0 1278 3 CESURF4
85 14 56.0 1536 1 XCHRP
86 14 56.0 2000 8 AF174649
87 14 56.0 2504 2 AC017284
88 14 56.0 2980 1 MNC1429B
89 14 56.0 2984 3 SEAJ4932
90 14 56.0 3519 3 AF222744
91 14 56.0 3850 6 E12103
92 14 56.0 3850 6 PPRC1GEN
93 14 56.0 3887 1 ECCR1DA
94 14 56.0 3895 1 BACTP1MEN
95 14 56.0 4223 8 ANCT7
96 14 56.0 4298 10 MCHMR
97 14 56.0 4349 1 ECOTRENC
98 14 56.0 4797 1 PS004358
99 14 56.0 4812 3 DR0DNA8PA
100 14 56.0 4851 1 AF123718

```

ALIGNMENTS

```

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS 1 (bases 1 to 1041)
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
source Location/Qualifiers
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 1041;
Best local Similarity 100.0%; Pred. No. 0.00022;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 gcatgcgcagcatcgcctcgt 25
Db 90 GCATGCCAGCATTCGATCTCGT 114

RESULT 2
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.

```

```

ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
source Location/Qualifiers
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 1041;
Best local Similarity 100.0%; Pred. No. 0.00022;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 gcatgcgcagcatcgcctcgt 25
Db 113 GCATGCCAGCATTCGATCTCGT 137

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION B10B gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
AUTHORS
JOURNAL Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES
source Location/Qualifiers
1. 1084
/organism="Escherichia coli"
/db_xref="taxon:562"
24..1064
/feature="B10B"
24..1064
/feature="B10B"
/codon_start=1
/protein_id="CAA00965.1"
/db_xref="GI:490219"
/db_xref="SWISS-PROT:P12996"
/translation="MAHPRMTLSQVTELEFEPKPLDLLEAAQVYRHFDPFQVST
LSTITGACPEDECKYCPSSRYKIGLEERLEMEVYVESARKAKAAGSTFCGAM
KNPHERDMPYLEOMVGKAMGLEACMTLGTLSOAOQLANAGLDYVNNHLDSPF
VGNITRTYQVROBLDTLEKVRDAGIKVSGGIVGEGVXRAGLLIOLANLTPPES
VFINLVKVGKPTPLADNDVDAPFIRIIVARIIMPSYRSLSGAEONNEOTQAMC
FMAGNSITFCGKILLTPNPEEDKDQLFRKLGLNPOQTAVLAGDNEQQRLEQALMT
PDTPRYNAAL"
BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 1084;
Best local Similarity 100.0%; Pred. No. 0.00022;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 gcatgcgcagcatcgcctcgt 25
Db 113 GCATGCCAGCATTCGATCTCGT 137

RESULT 4
LOCUS E00893 1121 bp DNA
DEFINITION Genomic DNA encoding biotin Synthetase.
E00893

```

ACCESSION E00893
 VERSION 1 GI:2169154
 KEYWORDS JP 1986149091-A/1.
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
 REFERENCE 1 (bases 1 to 1121)
 AUTHORS Hirono,Y., Kojima,T. and Kimura,H.
 TITLE DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND PRODUCTION OF BIOTIN
 JOURNAL Patent: JP 1986149091-A 1 07-JUL-1986;
 NIPPON SODA CO LTD
 COMMENT OS Escherichia coli
 PN JP 1986149091-A/1
 PD 07-JUL-1986
 PF 24-DEC-1984 JP 1984272605
 PI HIROMO YOSHITAKO, KOJIMA TAKAKAZU, KIMURA HIROSHI PC
 C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12N1:19),(C12P13/18, PC
 C12N1:19);
 CC strandedness: Double;
 CC topology: linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC *source: strain=Escherichia coli Nsl01;
 CC feature is identified by experimental;
 FH key Location/Qualifiers
 FT CDS 42..1079
 FT /product="biotin synthetase".
 FEATURES
 source location/Qualifiers
 1..1121
 /organism="Escherichia coli"
 /db_xref="taxon:562"
 BASE COUNT 289 a 296 c 325 g 211 t
 ORIGIN
 Query Match 100.0%; Score 25; DB 6; Length 1121:
 Best Local Similarity 100.0%; Pred. NO. 0.00022;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

source
 1..5793
 /organism="Escherichia coli"
 /strain="K-12"
 /db_xref="taxon:562"
 complement(96..574)
 /note="ORF 1"
 /codon_start=1
 /transl_table=1
 /protein_id="AA23513.1"
 /db_xref="GI:455168"
 /translation="MKLISNDLDCGKLEPHRYVNGMGYDGNISPHLAMDVPAGTK
 SEVAVTCYDPAPTGSGMHWVYVNLIPADTRVLPQGSGLVAPDGVLIATRDCKTG
 YDCAAPKGETHRYFTVHALDIERIDVDEGASAAVNGVHVFHSLOSTRAMS"
 complement(633..1925)
 /gene="bioA"
 complement(633..1925)
 /gene="bioA"
 /codon_start=1
 /transl_table=1
 /protein_id="AA23514.1"
 /db_xref="GI:457106"
 /translation="MTTDDLAFPQRIHTMPTSMTSPLPYVYVSAEGGELLISDGR
 LVGMSWMAIHGVNHPQLNAAMKSDIDAMSVMRGCTIHARATLCKRYAMSGRN
 ALBCEVPLADGSAVAEYAMKALQYWDKGEARQRLTERNGTIGDTFGMSYCDPDN
 SMHLMKGYLPENLFAPAPQSRMDGEMDERDNGFARLMAHHEIAAVITEPTVGA
 GMPMYHPENLKRIRKICDREGILLIDELATGFGTKLACEHAEIADILGCKA
 LTGGTMTLSATLTTRVAVERTISNGEACGFMHGTFMGNPLACAANAALIESGDMQ
 QGVADIEVQAREOLAPARDAEMVADVIVGALIGVETTHPVNMAALOKFVEGVNIR
 PFGKLIYMPETYLIPQOLRIIRAANVAVDEFFRQ"
 2012..3052
 /gene="bioB"
 2012..3052
 /gene="bioB"
 /codon_start=1
 /transl_table=1
 /product="biotin synthetase"
 /protein_id="AA23515.1"
 /db_xref="GI:145425"
 /translation="MAHRPWTLSQVTELEFEPKPLDLLFENQVYHROHPDPROYST
 ILSTKACGPEDCCKYCQTSRYKRTGLEERMEYEOYESARKAKAGSTRFCGAM
 KNHPRDMPYLEQVQGYCKAMGLPACMTLGTLSOARLANGLDYVNNHNDTPPE
 YGNITLTVYQERLDTEKVDAGIKVCSGIVGLGETVYKDRAGLLQLANLTPPES
 VPINLVKVGKPLADNDVDADFETITIAVARIMPTSVRLSAGRONMBOTAMC
 FMAGANSIFCYCKILLTTPNEPDKLDLQIFRKLGNIPQOTAVLADGNDQORLEOALMT
 PPTDEYVNAAL"
 3049..4203
 /gene="bioF"
 3049..4203
 /gene="bioF"
 /codon_start=1
 /transl_table=1
 /product="7-keto-8-amino pelargonic acid synthetase"
 /protein_id="AA23516.1"
 /db_xref="GI:145426"
 /translation="MSWQEKINAIADARRADALRRYPVAVQAGCRMLVADRDYLN
 SNDYLGLSHNPQIRAWQGAEGFGISGSGVSVVQALDEERLAWDLYSRA
 LLFISGANQAVIAAMAKEDRIADRLSHASILEASLSPSLRRFANDVYHRLR
 LLASPCGOVYVEGVSMQDSAPLAEIOOVQOIHGMMLMDVDAHGTGVIQGGG
 SCMDKVKPELLVYTFKGFPGVGAVALCSSVADYLILOFARHLIYSTMPADAOAL
 RSLAVINSDECDARREKLALITFRFRAGYODIPPTLADSCSAIOPILVGNSSALDL
 AEKLHQQGWVATIRPVPVAGTARLRLITRAHEMODIDRLLEYLHGNG"
 4190..4945
 /gene="bioC"
 4190..4945
 /gene="bioC"
 /codon_start=1
 /transl_table=1
 /protein_id="AA23517.1"
 /db_xref="GI:145427"
 /translation="MATVKKQAIAAAFGRAAHYEQHADLQROSADALLAMLPORRYT
 HVLDAGCGGMSRHMRERHAGVTAFLDISPPLVQARKDADADHDIADISPLATA
 TFDLMSNLAVQWCGNLSTALRELKRVYRPRKGVAVFTLLVQCSLPERHQVQVADERP

stem_loop 5583..5605
terminator 5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"

BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%: Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00022;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatcgccagcatcttcgactccgct 25
|||||
Db 206 GCATCGCCAGCATTTGATCCTCGT 230

RESULT 7
A38251 5872 bp DNA PAT 05-MAR-1997
LOCUS Sequence 6 from Patent WO9408023.
ACCESSION A38251
VERSION A38251.1 GI:2294849

KEYWORDS Escherichia coli.
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 5872)

AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL Patent: WO 9408023-A 6 14-APR-1994;
LONZA AG (CH)

COMMENT

Other publication PL 308301 950724
Other publication CA 2145400 94...1
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 9500809 950913
Other publication FI 951547 950331
Other publication JP 8501694T 960227.
location/Qualifiers

FEATURES
source 1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAA02329.1"
/db_xref="GI:2294850"
/translation="MSWQKINALDARRADALRRRYPVAGSGRWLVADDDROYLNF
SSNDYLSLHHPQITIRANOCAGDFGSGSGSHVSGSYVHOALEELAEELAEWLYGYSRA
LUTSGFRANQAVYIAMMAKEDRTAARLRLSHASLLEPASLSPSOLPRFAINDVTHLAR
LLASPCGQGVVTEGVFSMDGSAPLAEIQVYTOQNGMLWDDAAGTGVTEQCGK
SCSLQKVKPELLVTFGKFGVSGAAVLCSTVADYLLDFARHLIYSTWPAQOAL
RASLAVIRSDGDAAREKLAALITFRAGVODLPFTLLDSCSAIQPLIVDNGSRALQI
AEKLRQGCWVTAIRPTVPAGTARLRLTLAAHEMODIRLLLEVLHGNC"
3030..3045
/standard_name="BIOF RBS"
3043..3753
/gene="BIOF"

CDS 3043..3753
/gene="BIOF"
/standard_name="DEPHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="PFB SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYFVYCTDTFEVKTVASCALLQAAKACRTYRGYRPVAGS
EKTPREGIRNSDALALQNSSLQIDYAVNPNYFAPTSPHIIISAQGRPIESLVSAG
LRALEQOADVWLVGAGAGWFTPLSDFTFADWVTOPLVYIIVAGVKCINHAMLTA
OYIHNGLITAGVANDVPPGKRHAIEYMTTLIRMIAPLLGEIPMLAENPEAATGK
YINLAFYDASTLGFSTRL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%: Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00022;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatcgccagcatcttcgactccgct 25
|||||
Db 206 GCATCGCCAGCATTTGATCCTCGT 230

RESULT 8

A93674 5872 bp DNA PAT 22-JAN-2000
LOCUS Sequence 1 from Patent EP0798384.
ACCESSION A93674
VERSION A93674.1 GI:6741862

KEYWORDS Escherichia coli.
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O. and Brass, J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 1 01-OCT-1997;
LONZA AG (CH)

FEATURES
source 1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A-15/9"
1..96
/function="PROMOTER PTAC"
/evidence=experimental
23..28
45..49
/standard_name="PROMOTER PTAC"
/evidence=experimental
105..119
/standard_name="BIOB RBS NO.9"
/evidence=experimental
117..1157
/gene="BIOB"
117..1157
/gene="BIOB"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRPRMWTLSQVTELFKPLDLLFEAQGVHROHFDPRQOVST
LISIKTGACPEQCKVCPOSSRYKKTGLEARLMEVQVLESARKAKAAGSTRCCMAAW

promoter

-35.signal
-10.signal

RBS

gene

CDS

KNPHERDMPELEOMVGYKAMCELCMTLGTISESOAQLANAGLDYNNHNLDTSEPH
 VCNLTITRTYORLDTLEKVRDAGIKVGGGIVGGEVTKDAGLLDLANLPTPEPS
 VPIINLVKKGTPADNDVDADFLRTIARIMPTISYRLSAGRDQMDQDAMC
 FMAGANSIFVGGKLLTPNPEDDKDLQLEFRKLGLPDQTAVALAGNEDQORLEQALMT
 PDIDEYNNAAAL"

RBS
 2284..2297
 /standard_name="BIOC_RBS"
 2295..3050
 /gene="BIOC"

gene
 2295..3050
 /gene="BIOC"

CDS
 /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /product="PROTEIN"
 /protein_id="CAB65991.1"
 /db_xref="GI:6741864"

/translation="MATVAKQAAIAAFGRANAHYEQHDLQROSDALLMLPQRYT
 HYLDAGCGGMSRHRERHQAQVATLDSLPPMLVOARQKADAHYLAGDIESPLATA
 TFDLMSNLAVOMCGNLSTALRELYRVRPFGVAFYTLVGGSLPELHOAMQAVDERP
 HANRFLPPDEIPOSINGVHOHIOPTITLMDDDAISAKRSIKGIGATHLHGRDPRIL
 TRSOLQRLQIAMPOQCGRPLTYHFLFGLVIARE"

3742..3752
 /standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
 OXONONANOATE AMINOTRANSF."

gene
 3750..5039
 /gene="BIOA"

CDS
 3750..5039
 /gene="BIOA"

/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
 OXONONANOATE AMINOTRANSF."

EC_number="2.6.1.62"
 /codon_start=1
 /transl_table=11
 /evidence=experimental
 /product="DAPA SYNTHASE"
 /protein_id="CAB65992.1"
 /db_xref="GI:6741865"

/translation="MTTDDLAFDQRHITMPTYSPLPYVVSAGCELLISDCKR
 LVDMSWMAAIGHVNHPOLANAKSOIDAMSHVFGCITHAPALIELCKLVAMTPP
 LECVFIADSGVAEAVAMKALLOYWQAKGEARORELTFERNAGHDDPTGMSVCDPNS
 MSLIMKGYLPENLPAPAPQSMQDQEMDERDHWGPARLMAHRIIAVITTEPIVOCAG
 GMRMYHPBMLRIKIKICDREGILLIADETATGRTKLRACEHAETIAPDILCKAL
 TGTGMITLSATLTTPREVATTSNGAGCFHMGPTTPMGNPILCAANASLALIESGDWQ
 OVADIEVQLRQALPARDAEVADVRLGAIQVETTHPVNMAALOKFEVEQGVMIKP
 FGLTILMPPTIILPQOLRLTAAVNRAVODETFECQ"

RBS
 5088..5100
 /standard_name="ORF1_RBS"

gene
 5098..5574
 /gene="ORF1"

CDS
 5098..5574
 /gene="ORF1"

/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /evidence=experimental
 /product="PROTEIN"
 /protein_id="CAB69593.1"
 /db_xref="GI:6741866"

/translation="MKLISNDLRDQDKLPHRHVFNMGVYGDNIISPHLAMDVPAGIK
 SPVVTCTYDPAPTGSGMMHVVNLPADPRVLPQGRSGILVAMPDGVLOTFRDEGTG
 YDGAAPKGETHYIFTVAHLDIERIDVEGASGAVGVNHFHSLASASTAMES"

5583..5605
 /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
 TERMINATOR"

BASE COUNT
 1318 a 1552 c 1695 g 1307 t

Query Match
 Best Local Similarity 100.0%; Score 25; DB 6; Length 5872;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatgccagcatctgcactcgt 25
 Db 206 GCATGCCAGCATTTGCATCTCCTCGT 230

RESULT 9
 A93679
 LOCUS A93679 5872 bp DNA
 DEFINITION Sequence 6 from Patent EP0798384.
 ACCESSION A93679
 VERSION A93679.1 GI:6741867
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli.
 Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE
 1 (bases 1 to 5872)
 AUTHORS Birch,O. and Brass,J.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
 LONZA AG (CH)

FEATURES
 source
 1..5872
 /organism="Escherichia coli"
 /strain="DSM498"
 /db_xref="taxon:562"
 /clone="PBO30A15-9"
 1141..1156
 /standard_name="BIOF_RBS"
 1154..2308
 /gene="BIOF"

gene
 1154..2308
 /gene="BIOF"

CDS
 /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
 /EC_number="2.3.1.47"
 /codon_start=1
 /transl_table=11
 /evidence=experimental
 /product="KAPA SYNTHASE"
 /protein_id="CAB65994.1"
 /db_xref="GI:6741868"

/translation="MSWQEKINAAIDARADALRRRYPVVAGAGRILVADROLYN
 LFTSGFANQAVVIANMAKERIADRLSHASILEASLSPSLRRAHADVHLAR
 ILASPCPGQGVNTEGVFSMDGSDAPLAEIQQVYQNHGMVMDAHGTGVIQGGGK
 SWIQKVAPELIVTFEGKFGVGSAAVLCSSIVADYLLQFARHLIYTSMPDAOAL
 RASLAVIRSDGCDARREKIALITFRFAGVODLPPTLADSCSAIOPITVGNRAQL
 AEKLRQCGCWTAIRPPTVPAGTARLRLTLTAHEMODIDRLLEVLHNG"

RBS
 3030..3045
 /standard_name="BIOF_RBS"

gene
 3043..3753
 /gene="BIOF"

CDS
 3043..3753
 /gene="BIOF"

/standard_name="DETHIOBIOTIN SYNTHASE"
 /EC_number="6.3.3.3"
 /codon_start=1
 /transl_table=11
 /evidence=experimental
 /product="DTB SYNTHASE"
 /protein_id="CAB65995.1"
 /db_xref="GI:6741869"

/translation="MSKRYTYGTIDTEVGTVASCALLQAKAAGYTAAGYKPVASGS
 EKTPEGLNSDALALQRRSSLDQDYATVAVNPYTAEPISPHISAQEGRPTEISLVMSG
 LRALEQADWVLYEGAGGWFPPISDFTFPADWVYQEDLPVILVGVYAGCINAMILTA
 QVIOHAGITLAGWANDVTPGKRHAQVMTTITRMIPAPILGELPMLAENPENAAATGX
 YINLAFVDASTIGFTSRL"

BASE COUNT
 1318 a 1552 c 1695 g 1307 t

Query Match
 Best Local Similarity 100.0%; Score 25; DB 6; Length 5872;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

	Matches	25;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	gcacgcacgaattccatctcgt	25							
Db	206	GCATCGCCAGCATTTTCGATCCTCGT	230							
RESULT	10									
LOCUS	ARI01809	5872 bp	DNA		PAT		14-FEB-2001			
DEFINITION	Sequence 1 from patent US 6083712.									
ACCESSION	ARI01809									
VERSION	ARI01809.1	GI:12812607								
KEYWORDS										
SOURCE	Unknown.									
ORGANISM	Unclassified.									
REFERENCE	1 (bases 1 to 5872)									
AUTHORS	Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.									
TITLE	Biotechnological method of producing biotin									
JOURNAL	Patent: US 6083712-A 1 04-JUL-2000;									
FEATURES	Location/Qualifiers									
source	1..5872									
	/organism="unknown"									
BASE COUNT	1318 a 1552 c 1695 g 1307 t									
ORIGIN										
Query Match	100.0%;	Score 25;	DB 5;	Length 5872;						
Best Local Similarity	100.0%;	Pred. No. 0.00022;								
Matches	25;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;					
Qy	1	gcacgcacgaattccatctcgt	25							
Db	206	GCATCGCCAGCATTTTCGATCCTCGT	230							
RESULT	11									
LOCUS	ARI01810	5872 bp	DNA		PAT		14-FEB-2001			
DEFINITION	Sequence 6 from patent US 6083712.									
ACCESSION	ARI01810									
VERSION	ARI01810.1	GI:12812608								
KEYWORDS										
SOURCE	Unknown.									
ORGANISM	Unclassified.									
REFERENCE	1 (bases 1 to 5872)									
AUTHORS	Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.									
TITLE	Biotechnological method of producing biotin									
JOURNAL	Patent: US 6083712-A 6 04-JUL-2000;									
FEATURES	Location/Qualifiers									
source	1..5872									
	/organism="unknown"									
BASE COUNT	1318 a 1552 c 1695 g 1307 t									
ORIGIN										
Query Match	100.0%;	Score 25;	DB 6;	Length 5872;						
Best Local Similarity	100.0%;	Pred. No. 0.00022;								
Matches	25;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;					
Qy	1	gcacgcacgaattccatctcgt	25							
Db	206	GCATCGCCAGCATTTTCGATCCTCGT	230							
RESULT	12									
LOCUS	AE000180	11022 bp	DNA		BCT		01-DEC-2000			
DEFINITION	<i>Escherichia coli</i> K12 MG1655 section 70 of 400 of the complete genome									
ACCESSION	AE000180	U00096								

complement(147. .1430)
 /gene="ybhC"
 /function="orf; Not classified"
 /note="1427: 98 pct identical to fragment YBH_C_ECOLI
 SW:P46130 (300 aa) but contains 127 additional C-terminal
 residues"
 /codon_start=1
 /transl_table=11
 /product="putative pectinesterase"
 /protein_id="AAC73859.1"
 /db_xref="GI:1786989"
 /translation="MNTFSVRLALALAFGVYITACSSTPPDORPDTATPCTSSRI
 LSKEAONPDQHYEASTLPGAAANPSPITLPAQPDVGPAGTGGYTHITIAAVD
 AAIIRTKNRQYIAVPEYOGTYVPAAPGSLITLGTGKPKIDYKISLSDGKMSIA
 DRRHDVNRGKYMGPAMVYWDSCSKRSISGVLCSAVFMSQNNGLQJONLIENT
 LGSVDACNHPAVVALRTGDQVQINNVLIGKONFVFNAGVONRLETNRQPTLV
 NSYIEGDVDSISGCAVVDNTEFRVNSRTQEAIVFAPATLSNIYYGLAVNSRN
 AFDGVAOLGSLVDANTNGOYVIRDSAINEGFNITAKPMADAVISNRPFAINTGSVD
 DNDETQRMINDPTNRMWENNRGVSKVAEAKK"
 complement(1507. .1535)
 /note="factor Sigma70: predicted +1 start at 806574"
 complement(1582. .2058)
 /gene="ybhB"
 /note="b0773"
 complement(1582. .2058)
 /function="orf; Unknown"
 /note="1158: 99 pct identical to YBH_B_ECOLI SW: P12994"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC73860.1"
 /db_xref="GI:1786990"
 /translation="MKLISNDLRDGLKPHRHVNGCYGDNISPILANDIYPACIK
 SFVYCYDPAQSGSGMHWVYVNLPAQTRVLPGCGSGLYAMPDGLQTRDGGKIG
 YDGAAPKGETHRYITFVHALDRIYDDEGASGAVGVNHFSLASASITAMFS"
 complement(2108. .2124)
 /note="central position to predicted promoter:85.5"
 /bound_molecule="RhaS predicted site"
 complement(2117. .3406)
 /gene="bioA"
 /note="b0774"
 complement(2117. .3406)
 /gene="bioA"
 /EC_number="2.6.1.62"
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="1429: 100 pct identical to BIOA_ECOLI SW: P12995"
 /codon_start=1
 /transl_table=11
 /product="7,8-diaminopelargonic acid synthetase"
 /protein_id="AAC73861.1"
 /db_xref="GI:1786991"
 /translation="MTTDLAPDORHIMHPYTSWSPILVYVSAEGEELLISDGR
 LVGSSMAAIIHGYNHPOINAMKSOIDAMSVHMGITTHAIFELCKRLVAMPPTP
 LECVPLADSGSAVAVYAMKALDYOQAKGBARORLTFRNGYHGDFFGAMSVADPN
 MSLIMKGYLPENLFAPOSQKRDGDEWDFARLMAHREIATVLEPIVQAG
 GMRVYHPWLKRIKIDREGILLADIETATGRTGKIFACEHAETADPDIICGAL
 TGGITMLSAITLTREVAETISNGEACPMHGPFGMPNPLVCAANAASIALILSSGMO
 OVADIPEOLREOLAPADAEMVADVFLAIGVETTHVYMAALOKRFEVDEGVNIRP
 FCKLIYMPYITILPOOLRIITAAVRAVODEFFGQ"
 complement(2193. .2221)
 /gene="bioA"
 /note="factor Sigma70: predicted +1 start at 807260"
 3411. .3450
 /note="central position to bioB promoter: -20"
 /bound_molecule="BioB documented site"
 3411. .3450
 /note="central position to predicted promoter: -20; genetic
 evidence for the site"
 /bound_molecule="BioB predicted site"
 complement(3411. .3450)
 /note="central position to predicted promoter:50"

protein_bind
 /bound_molecule="BioB predicted site"
 complement(3411. .3450)
 /note="central position to bioA promoter:50"
 /bound_molecule="BioB documented site"
 3413. .3441
 /note="factor Sigma70; promoter bioB; documented +1 at
 808525"
 complement(3447. .3473)
 /note="factor Sigma70; promoter bioA; documented +1 at
 808515"
 3493. .4533
 /gene="bioB"
 /note="b0775"
 3493. .4533
 /gene="bioB"
 /EC_number="2.8.1.-"
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="0346: 99 pct identical to BIOB_ECOLI SW: P12996"
 /codon_start=1
 /transl_table=11
 /product="biotin synthesis, sulfur insertion?"
 /protein_id="AAC73862.1"
 /db_xref="GI:1786992"
 /translation="MAHRPRMTLSQVTELEKPLDILFEAQOVHROHFDPROGVST
 LLSIKTGACPEDECKYCOSSRYKTLGEMLERLAEVQVLEAKAKAAGSTRPCGAAM
 KNPHERMDPYLEGMVQVQKAMLEAGMTGTISESOAORLAQMDIYNNHNDPSPEF
 YGNIITTPRYOERDPLEKVRDAGIKVCGVIGVGEVYKPDAGILLOLANIPPEP
 VPIMLVVKYKGTPLIADNDVDPAFDFTIRAVRIMPPISYVLSGROMNBOYAMC
 PMAGANSIFVCGKLLTTPNPEDEKDLQFRKLGLNPQOAVLAGNEQOQRIEALMT
 PPTDEYNAAL"
 4530. .5684
 /gene="bioF"
 /note="b0776"
 4530. .5684
 /gene="bioF"
 /EC_number="2.3.1.47"
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="0384: 100 pct identical to BIOF_ECOLI SW: P12998"
 /codon_start=1
 /transl_table=11
 /product="8-amino-7-oxononanoate synthase"
 /protein_id="AAC73863.1"
 /db_xref="GI:1786993"
 /translation="MSMOEKINAALDARRADALRRYPVAGAGRMVLVADPROLYNF
 SSNDYLGSHHPOILIRAWQGAEOFGISGSGVSGVYHQALEBELAEMVGSRA
 LLFISGFAANOAVIAAMAKEDRIADRLSHASILFAASISPSQLRFAHNDYHLAR
 LLASPCGOAVVTEGVFSMDGDSAPLAETIOGVYQOINGMILVDAHGTGVTGOGKG
 SCWLQKVPPELLVYTFGKFGVSCAAVLCSTVADYLLQFARRHLIYSTSMPPAQAL"

Query Match 100.0%; Score 25; DB 1; Length 11022;
 Best local Similarity 100.0%; Pred. No. 0.00021;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 gcatgcagcatcgcacccctcgt 25
 |||||||||||||||||||||||||
 Db 3582 GCATGCCAGCATTTGATCCTCGT 3606

RESULT 13
 AEO05258 LOCUS 13501 bp DNA BCT 21-MAR-2001
 DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
 OF 155.
 ACCESSION AEO05258 AEO05174
 VERSION AEO05258.1 GI:12513751
 KEYWORDS
 SOURCE
 ORGANISM Escherichia coli O157:H7 EDL933.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE 1 (bases 1 to 13501)

Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 cagcatttcgacttcgt 25
|||||
Db 57 CAGCATTTTCGATCTCGT 74

RESULT 17
G01446 479 bp DNA STS 19-AUG-1999
LOCUS Dm0466 Drosophila P1 library Drosophila melanogaster STS genomic
DEFINITION clone DS07967 sp6, sequence tagged site.
G01446
ACCESSION G01446.1 GI:684849
VERSION G01446.1 GI:684849
KEYWORDS STS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 479)
Rubin, G.
Berkeley Drosophila Genome Project
Unpublished (1994)
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Contact: Berkeley Drosophila Genome Project
Primer A: CTCCTACTGGGACTTCCTA
Primer B: CACAAAGGCGATATCCGA
STS size: 192
PCR Profile:
Annealing: 58 degrees C PCR Cycles: 30
Protocol: P1 Library Pools
Template: 1 uM each
Primer: 250 uM each
dNTPs: 0.05 units/uL
Taq Poly: 15 uL
Total Vol: 15 uL

Buffer:
MgCl2: 1.5mM
KCl: 50 mM
Tris-HCl: 50 mM
pH: 8.3
Gelatin: .001 %

The P1 library has been distributed to 16 regional sites. A list of these sites is available from Flybase, via anonymous ftp to ftp.bio.indiana.edu in the file flybase/allied-data/genome-projects/1b1/UBL.doc.

FEATURES

Source
1. .479
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="DS07967"
/clone_id="Drosophila P1 library"
/note="Vector: Pax10Sact1; The P1 library was made by D. Smoller in D. Hartl's lab (see Smoller et al., Chromosoma 100: 487). Chromosomal position in the Hartl lab (see Hartl et al., PNAS 91: 6824). STS's were generated by sequencing the ends of the Drosophila insert in these P1 clones, by the W. Krummery, C. Martin, and M. Palazzolo lab at LBL."
STS primer_bind complement(233..250)
BASE COUNT 113 a 120 c 97 g 145 t 4 others
ORIGIN

Query Match 68.0%; Score 17; DB 11; Length 479;
Best Local Similarity 100.0%; Pred. No. 11;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 gcatgccagcatttcg 17
|||||
Db 396 GCATGCCAGCATTTTCG 412

RESULT 18
AE007981/c 10464 bp DNA BCT 14-AUG-2001
LOCUS AE007981 Agrobacterium tumefaciens strain C58 circular chromosome, section 39 of 254 of the complete sequence.
DEFINITION AE007981 AE007869
ACCESSION AE007981.1 GI:15155357
VERSION AE007981.1 GI:15155357
KEYWORDS Agrobacterium tumefaciens.
SOURCE Agrobacterium tumefaciens; alpha subdivision; Rhizobiaceae group; Bacteria; Proteobacteria; Rhizobium.
ORGANISM Rhizobiaceae; Rhizobium.
1 (bases 1 to 10464)
Hinkle, G., Slater, S.C. and Goodner, B.
Complete Genome Sequence of Agrobacterium tumefaciens C58 (Rhizobium radiobacter C58), the Causative Agent of Crown Gall Disease in Plants
Unpublished
2 (bases 1 to 10464)
Hinkle, G., Slater, S.C. and Goodner, B.
Direct Submission
Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street, Cambridge, MA 02139, USA
Location/Qualifiers
1..10464
/organism="Agrobacterium tumefaciens"
/strain="C58"
/db_xref="taxon:358"
complement(162..1067)
/gene="AGR_C_768"
complement(162..1067)
/gene="AGR_C_768"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_768p"
/protein_id="AAK86250.1"
/db_xref="GI:15155358"
/translation="MDLQTAVAETIAALTNNAIVQVPADAKCIPSTVALLD PTSGRLADYSAIPAPRTAPRRKQGTAAVETLESFQVVERQGTENSVIFAVTRDP TTTAVIDYHDDDDNGKHRVHYAPLSEEMKAMOSINKALITONEFAEIDHIAELIA SPDSDEKIDLELFRKVAVYVNEIYLSGQIINETRYKTAVKLQTESQIIVEEDH KNAAGDTTPYGVGVNLINAPFROGESIRLPVRLRLRREGKLSWTCMLRPDPHTTNA VTFALHETAAELGLPFAKPEMSA"
complement(1127..1459)
/gene="AGR_C_769"
complement(1127..1459)
/gene="AGR_C_769"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_769p"
/protein_id="AAK86251.1"
/db_xref="GI:15155359"
/translation="MKKIDHASPITIGLNGQLNPAFSAISIGTLEKLSDMSEANPMQ SFKGSVILKLSLXKDGKMTIADHESKTPKLPKKNVFMVYEDGALSTEHPRDHMF TPREVSAS"
complement(1490..1702)
/gene="AGR_C_771"
complement(1490..1702)
/gene="AGR_C_771"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_771p"
/protein_id="AAK86252.1"

gene
CDS
gene
CDS

RESULT 19
AC019940 53067 bp DNA HTG 03-JAN-2000
LOCUS AC019940/c
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, In ordered pieces.
ACCESSION AC019940
VERSION AC019940.1 GI:6664957
KEYWORDS HTG; PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 53067)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
COMMENT This sequence was identified as CDM-10211463 by the submitter. For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1. 53067
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 14944 a 11879 c 11720 g 14524 t
ORIGIN

Query Match 68.0%; Score 17; DB 2; Length 53067;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatccgcagcattcg 17
|||||
Db 13629 GCATCGCCAGCATTCG 13613

RESULT 20
AC007809 183439 bp DNA INV 24-FEB-2001
LOCUS AC007809
DEFINITION Drosophila melanogaster, chromosome 3R, region 88C-88C, BAC clone BACR45M03, complete sequence.
ACCESSION AC007809
VERSION AC007809.8 GI:13122709
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 183439)
AUTHORS Celisner, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C., Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A., Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorset, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferreira, S., Frise, E., Galle, R.F., Galt, N.S., George, R.A., Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Putl, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svitskas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
TITLE Sequencing of Drosophila chromosome 3R, region 88C-88C
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 183439)
AUTHORS Celisner, S.E., Agbayan, A., Arcaina, T.T., Baxter, E., Blazek, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,

Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotian, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Pak, S., Pfeiffer, B., Poon, L., Sequelra, A., Sehl, H., Shit, E., Svitskas, R.K., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT On Feb 24, 2001 this sequence version replaced gi:6623906. Sequence submitted by: Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121 Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.berkeley.edu.
FEATURES
source
1. 183439
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="88C-88C"
/clone="BACR45M03 (D718)"
/clone_id="RPC1-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBACE3.6)"
BASE COUNT 53300 a 38849 c 38854 g 52436 t
ORIGIN

Query Match 68.0%; Score 17; DB 3; Length 183439;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcgcagcagcattcgatc 20
|||||
Db 66707 TCGCAGCATTCGATC 66723

RESULT 21
AC092401 185200 bp DNA INV 04-JUL-2001
LOCUS AC092401
DEFINITION Drosophila melanogaster, chromosome X, region 13E-13F, BAC clone BACR25C18, complete sequence.
ACCESSION AC092401
VERSION AC092401.1 GI:14595766
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 185200)
AUTHORS Celisner, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C., Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A., Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorset, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferreira, S., Frise, E., Galle, R.F., Galt, N.S., George, R.A., Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Putl, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svitskas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
TITLE Sequencing of Drosophila chromosome X, region 13E-13F

JOURNAL
REFERENCE
AUTHORS
Unpublished
2 (bases 1 to 185200)
Celniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanalides, P.G., Brannon, R.C., Rogers, J., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A., Carlson, J.W., Center, A., Chapple, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorselt, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferriera, S., Frisoe, E., Galle, R.F., Garg, N.S., George, R.A., Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattel, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Piltman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svirska, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
Direct Submission
Submitted (04-JUL-2001) Berkeley Drosophila Genome Project, MS 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, US
COMMENT
Sequence submitted by:
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.berkeley.edu.
FEATURES
source
1. 185200
/organism="Drosophila melanogaster"
/strain="Y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="X"
/map="13E-13F"
/clone="BACR25C18 (D1403)"
/clone_lib="RPC1-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBACE3.6)"
BASE COUNT
52552 a 40775 c 40327 g 51546 t
ORIGIN

Query Match 68.0%; Score 17; DB 3; Length 185200;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcatctcg 17
|||||
Db 88505 GCATGCCAGCATTTTCG 88521

RESULT 22
AC013431 191558 bp DNA INV 31-MAY-2001
LOCUS
DEFINITION
BACR22H11, complete sequence.
AC013431
VERSION
AC013431.9 GI:14269669
KEYWORDS
HTG.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 191558)
Celniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H., Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A., Carlson, J.W., Center, A., Chapple, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorselt, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferriera, S., Frisoe, E., Galle, R.F., Garg, N.S., George, R.A.,

JOURNAL
REFERENCE
AUTHORS
Unpublished
2 (bases 1 to 191558)
Celniker, S.E., Agbayani, A., Arcalini, T.T., Baxter, E., Blazej, R.G., Burenhoff, C., Chapple, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hinkle, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Lee, B., Lewis, S., Li, P., Ling, H., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Richards, S., Sethi, H., Svirska, R.R., Wan, K.H., Webster, D., Woolley, P., Yang, S., Yee, M., Yu, C. and Rubin, G.M.
Direct Submission
Submitted (11-NOV-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
On May 31, 2001 this sequence version replaced g1:6838815.
COMMENT
Sequence submitted by:
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.berkeley.edu.
FEATURES
source
1. 191558
/organism="Drosophila melanogaster"
/strain="Y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="X"
/map="13E-13F"
/clone="BACR22H11 (D1191)"
/clone_lib="RPC1-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBACE3.6)"
BASE COUNT
54283 a 41835 c 41651 g 53789 t
ORIGIN

Query Match 68.0%; Score 17; DB 3; Length 191558;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcatctcg 17
|||||
Db 67235 GCATGCCAGCATTTTCG 67251

RESULT 23
AE003705 232744 bp DNA INV 05-OCT-2000
LOCUS
DEFINITION
Drosophila melanogaster genomic scaffold 142000013386035 section 30 of 105, complete sequence.
AE003705 AE002708
VERSION
AE003705.1 GI:7299886
KEYWORDS
HTG.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 232744)
Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanalides, P.G., Scheeler, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,


```

KSPRLSNSSASSLPPCRRHMLDDEGLFMEILIRVAMFLIDIVLCHTILVNGVLI
HADGSLFLAKGPRNKYIVAKLFDVYORTALKADAVTASAEITILPFGIAGVAQ
TKOMINIKENKARNCETIDLTGKTNAIILCPONCEGDIIGVAQIINKNGME
EDHDEVEIFRRYIFPGIGIONQLEFMSVOERRNOIILINARSIFEEONNLECLT
KILTEARELLKERCSPFLVLDCEASHLEKIIEKPNOPATRAIKASPEEKRRN
RFLVFELEGEOYANVRSPVSEISSSTIAOIAQVATTCOTVINICDVTENVRHNO
IRAEDEIDSTOALICMPIMNAOKKIVIGVQOLINKAGVPTSPDSIFEPAPAECCIG
IHNTOMYENACKIMAKOKVALRELSYHAGVQOLINAKGVPSTSPQPOASGSMCR
FELVDDTCRAVIRMPNOCNIVSOPITPDVLCRWVLSRKNVPRVKHNHMLNVA
OTMFAMLKTKGKMERFMTDLEIGLIVACCHLDHRTNNAEOTKTESPLALITYST
MEHHHDHRLAFLDLRLAFLRDLARLDELVDLVEELVDPITLFLVEYERLIR
TTLFELDELDEEDVDVSVVAVLVDLSVLLSAALMAVTKTIERKT
RNFEMDALSPEDRSVKTESALISTLAWFKRNKFAELVDEPOGDEKLOITOPAMDRKRD
LTCGMWACIVSALAKPWEVOHKVAKLVADEPOGDEKLOITOPAMDRKRD
ELPKMVGIFDVICLPLRVLCDFPMTPLPTEGLEENRNQDLAEVEMGLTIDH
DTRDKVEEPKACADEIKDIEFTVTLTLCNCOOSHSHTPHORNGSKLSMKT
GALGRKVRKSLKTLVNSMDSPKTSKLSHSEMDDKSPSPQPOASGSMCR
MSASSTSSAGQMDKRSKRLCALL
Join(40359..40547,40626..40826,40891..41076,41138..41210,
41284..41410,41461..42034,42102..42262)
/gene="Cyp313a1"
/product="CT11271"
/db_xref="FLYBASE:FBan0003360"
/db_xref="FLYBASE:FBgn0038236"
<40359..>42262
/gene="Cyp313a1"
/note="CG3360"
/map="88C5-88C6"
/db_xref="FLYBASE:FBan0003360"
/db_xref="FLYBASE:FBgn0038236"
Join(40391..40547,40626..40826,40891..41076,41138..41210,
41284..41410,41461..42034,42102..42262)
/gene="Cyp313a1"
/note="Cyp313a1 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0003360"
/db_xref="FLYBASE:FBgn0038236"
/protein_id="AAF5065.2"
/db_xref="GI:10726530"

Query Match      68.0%; Score 17; DB 3; Length 232744;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcgccagattcgatc 20
|||||
Db 106438 TCGCCAGCATTTGCATC 106454

RESULT 24
AC017740/c
LOCUS
DEFINITION
AC017740 237119 bp DNA 10-DEC-1999
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
AC017740
VERSION
AC017740.1 GI:6554259
KEYWORDS
HTG; HTGS; PHASE2.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 237119)
Adams,M. and Venter,J.C.
Direct Submission
Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10211941 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
Location/Qualifiers
1..237119
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 67689 a 49909 c 50172 g 69349 t
ORIGIN

Query Match      68.0%; Score 17; DB 2; Length 237119;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcgccagattcgatc 20
|||||
Db 103681 TCGCCAGCATTTGCATC 103665

RESULT 25
AE003500
LOCUS
DEFINITION
AE003500 327446 bp DNA INV 04-OCT-2000
Drosophila melanogaster genomic scaffold 142000013386053 section 17
of 30, complete sequence.
ACCESSION
AE003500 AE002593
VERSION
AE003500.2 GI:10728273
KEYWORDS
HTG.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 327446)
Adams,M.D., Geinlinger,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Mananides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazek,R.G., Champs,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,F.G., Heil,G., Nelson,C.R., Gabor
Miklos,G.L., Abrell,J.F., Appayani,A., An,H.J.,
Andrews,Plannkocn,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,D., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkov,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brotlier,P., Butts,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
Fleischmann,W., Foster,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Gelpart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Jbergwam,C., Jalali,M., Kalush,F., Kodira,C.D., Kraft,C.,
Kremetz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,S.M., Murphy,B., Murphy,L.,
Nusken,D.R., Pacleb,J.M., Palazolo,M., Pitman,G.S., Pan,S.,
Pollard,J., Puri,Y., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shew,B.C., Siden-Klamos,I.,
Simpton,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svrtk,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
Weinstock,G.M., Wellsteinbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yen,R.F.,
Zaveri,J.S., Zhao,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,J.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006
1 (bases 1 to 327446)
TITLE
JOURNAL
MEDLINE
REFERENCE

```


Query Match	66.0%;	Score 17;	DB 3;	Length 327446;
Best Local Similarity	100.0%;	Pred. No. 10;		
Matches	17;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	gcacgcacgacattccg	17	
Db	70808	GCATGCCAGCATTTTCG	70824	
RESULT 26				
LOCUS	AF146614/c			
DEFINITION	AF146614	1186 bp	DNA	ECT 09-AUG-1999
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
FEATURES				
source				
gene				
CDS				
gene				
CDS				
sig_peptide				
CDS				
Query Match	64.0%;	Score 16;	DB 1;	Length 1186;
Best Local Similarity	100.0%;	Pred. No. 42;		
Matches	16;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	gcacgcacgacattcc	16	
Db	993	GCATGCCAGCATTTTC	978	
RESULT 27				
LOCUS	AF206717/c			
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
FEATURES				
source				
mat_peptide				
misc_feature				
gene				
CDS				
BASE COUNT	295 a	315 c	323 g	253 t
ORIGIN				
Query Match	64.0%;	Score 16;	DB 1;	Length 1186;
Best Local Similarity	100.0%;	Pred. No. 42;		
Matches	16;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	gcacgcacgacattcc	16	
Db	993	GCATGCCAGCATTTTC	978	
RESULT 27				
LOCUS	AF206717/c			
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
FEATURES				
source				
SwissProt Accession Numbers P39691 and P21892"				
/codon_start=1				
/transl_table=11				
/product="DsbC precursor"				
/protein_id="AAD47610.1"				
/db_xref="GI:5712696"				
/translation="MKRGILLSLVATVATNEFAHDDAAIKOTLHAGSYARGDPAPFA				
DGRMKTIVITDSGLVYISDEGQAFAPARALRPRTPVNTTKILIGKDALOEOMIVY				
KAQEKHVIITVEITISGCGYCHLHDOMDYNAIGITVYRLAFPPQGMKSPARKMOSI				
WCYADNRKAFDSAMKGDATSAATCTDIAAHQOLQIQCGVGTATVQLDGLVYPGTLQ				
GPKEMLAMLEFAHKASKRTKG"				
378..1028				
/gene="dsbC"				
/product="DsbC"				
669..680				
/gene="dsbC"				
/note="Region: putative CPHC motif typical of the active				
sites of disulfide isomerases"				
1044..21186				
/gene="recJ"				
1044..21186				
/gene="recJ"				
/note="theoretical protein; similar to Erwinia				
chrysanthemi and Escherichia coli RecJ single-stranded				
DNA-specific exonucleases in SwissProt Accession Numbers				
P39693 and P21893"				
/codon_start=1				
/transl_table=11				
/evidence="not_experimental				
/product="RecJ"				
/protein_id="AAD47612.1"				
/db_xref="GI:5712698"				
/translation="MITQLRRRPLAEDIDLPTVPPILRLRYAHGCVKKAQNELERLSL				
GLL"				

```
/organism="Shuttle vector p13"
/specific_host="Deinococcus radiodurans SARK"
/db_xref="taxon:115340"
/note="derived from pue10"
105..593
/contig="OrfA"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="AA04052.1"
/db_xref="GI:7263004"
/translation="MGQOLVROGCDVPPNNGOQDVVRLPILPGVKKKFAALRG
GKERQNDPGDISMRKMTBEAFHPPDGGGVRASGEARCEFSMPDVLAREGDHDS
EOLDLVLELREVRGEAGQIGODIGRRVPSRVSRTFSSPAGNCRVLRIMGMOKRER
AV"
misc_feature
complement(257..683)
/note="5 mismatches with sequence deposited in Genbank
Accession Number M94966"
649..653
/contig="OrfB"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="AA04047.1"
/db_xref="GI:7262999"
/translation="MRFYVAFDGSSEPSORALKAQITLKGAGALEVTVLEPPVSG
FYWAGLNDPDYREMLFONARQAOATQTLIGACVGGVQVTLGRPIEVLTEAHGV
DLVVGTHGYRGMDRIILGCVETILRRAEVLLVVR"
1187..2436
/contig="HtrU"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="HtrU"
/protein_id="AA04048.1"
/db_xref="GI:7263000"
/translation="MKTAMPILTLILLAGCTDPTESRTQTSSTRTATTNARGA
ARSSSPSGAGOTASPDGRLVDGNTIDVYKDDGVVTRFDQSGGALFESP
LAPDAPBDEBPTGSGSFLIDRECHILITNVYIODATDIRLHONRDYASVYG
TAPAYDLALIRASGARPPGAPAGGORSAGCRKDCIAPGLFTYTCGILITAVK
RVIPMGVESTIPONSQVTDAAINPNSGGLVNSRGEVGVNTOILSPAGAVGQNA
GVGFAPVNVVKSILPRLRAGEITVPRIGIVSNLQALTPSARELGIPEDGVLOS
VERGTPAAAGLGRGPRSORPDGDIRLGDLITAVDGOEVSTVODLQGVLLGKOSD
EVLTILRROGQTLQRKLTLS"
3113..3118
3127..3678
/note="OrfC: proline-rich protein"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="AA04053.1"
/db_xref="GI:7263005"
/translation="MTFHCAAMASARSTRSQSCAATHSRSLGVGESAAVPAATV
ARLPVPVPRASAPFLGLSDMHPRVPSGAAPLPVVPVPSRVGSCSPAVPLSKLP
RPSSTPHIKISFCWGRCSFSPDLLPKQIVVGRAGAGPPLRPHFWPARPACFPKP
AARPLVSLISCIICAVTVLLSGI"
3691..4038
/note="OrfD"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="AA04054.1"
/db_xref="GI:7263006"
/translation="MTTTPAPLFSASAVLGSNPSVAVRSGCGAAMWCMNRHTAP
ALALRTPAASAPSTYATVAKMYAQRGRLRVQVRRAPCGVLMKCSGPRPDLAAGQ
```

```
4090..4551
/note="OrfE"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="AA04055.1"
/db_xref="GI:7263007"
/translation="MAAPTRKHTSPPPMPLIAGLMSRVAPVALTPPPLPPPCSAV
TPPPACACFPLVLRPTSPPPGSSSTPPPSPCFPSPGSAAPRRMPCGMPSSAA
CPGFSLAPLRPPRPPAGRPPLQAFPAAGGRRRPCACSSFPAPSRLSFL"
4666..4670
4678..5364
/note="OrfF"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="AA04056.1"
/db_xref="GI:7263008"
/translation="MSELTFRSLSPDHARALVALGADHAPALARAADLVMOQLSGP
APVTDARRSRAPARTLPLPGVDYGRWDSAPLPTARACKLNLTAAGAADIGMAR
ALTRGDIYAAEQITCEGMQLOITONHRRQAALAAQQAQOAPAPALTRTORALHK
OLAGIWTEDERTYQAAKIRRPVESITLALNODERAAALLVAKAEQARRAVALAOD
PAAYGALLR"
5407..5973
/contig="ardU"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="ardU"
/protein_id="AA04049.1"
/db_xref="GI:7263001"
/translation="MTYTHPLIERHPDAPALWICGLAATNACKLHCAMQASSDTRAE
MFGAIEELIKASPEPHAEEMDITMDNMPFAGRTLDSAAATVAAALDALSRADELY
AAWEMRGAEEMDADKITDAYLGRFDSVEDYAAQYLDSDGALQVEPMLRPIYNTAAL
GRDMEINDVYEGKNKGHFNGHA"
6148..6153
6164..6445
/note="OrfG"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="AA04057.1"
/db_xref="GI:7263009"
/translation="MQVARDMPKLPPEVAAEDVLTVKLVLYITMANPGEHSDSLNA
ELGVRPOGAIKRLERGVLYOEAPRGPKPKYRAQTKAKRPAQEEPMP"
6736..6741
6748..7749
/note="OrfH: alanine-rich protein"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="AA04058.1"
/db_xref="GI:7263010"
/translation="MOKITFAAIIIFSAALATACGGGPPNTPPAPAPASADVADPTDG
AVLASGTRKMSAQAVOSLGRDLATAPATPAPVAVYVKKGPGSGRYAGOKRAFLG
WKISGKAPPAHLTGATSGSYALATLPGADVKYTTGILSVSLPLASNLNMTSRSRT
SYAAVYGGTCVAASSALTLANGSAAEFPGDAPLVVCEGVYVAFRTNPAPATLELD
WVQSSFTTTSAGRGALVVSFAQVDPKAVMLMKRYOQAAEEGADDAVAVASGEMOT
LSSPANTRRFPAVVDQKMVQEHYGLAVQAFKDBGTFHGELTLREGLLCEPGATPS"
7953..9159
/contig="repu"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="repu"
/protein_id="AA04059.1"
/db_xref="GI:7263011"
/translation="MOKITFAAIIIFSAALATACGGGPPNTPPAPAPASADVADPTDG
AVLASGTRKMSAQAVOSLGRDLATAPATPAPVAVYVKKGPGSGRYAGOKRAFLG
WKISGKAPPAHLTGATSGSYALATLPGADVKYTTGILSVSLPLASNLNMTSRSRT
SYAAVYGGTCVAASSALTLANGSAAEFPGDAPLVVCEGVYVAFRTNPAPATLELD
WVQSSFTTTSAGRGALVVSFAQVDPKAVMLMKRYOQAAEEGADDAVAVASGEMOT
LSSPANTRRFPAVVDQKMVQEHYGLAVQAFKDBGTFHGELTLREGLLCEPGATPS"
7953..9159
/contig="repu"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="repu"
/protein_id="AA04059.1"
/db_xref="GI:7263011"
/translation="MOKITFAAIIIFSAALATACGGGPPNTPPAPAPASADVADPTDG
AVLASGTRKMSAQAVOSLGRDLATAPATPAPVAVYVKKGPGSGRYAGOKRAFLG
WKISGKAPPAHLTGATSGSYALATLPGADVKYTTGILSVSLPLASNLNMTSRSRT
SYAAVYGGTCVAASSALTLANGSAAEFPGDAPLVVCEGVYVAFRTNPAPATLELD
WVQSSFTTTSAGRGALVVSFAQVDPKAVMLMKRYOQAAEEGADDAVAVASGEMOT
LSSPANTRRFPAVVDQKMVQEHYGLAVQAFKDBGTFHGELTLREGLLCEPGATPS"
```

```

/function="initiation of plasmid replication"
/codon_start=1
/transl_table=11
/product="RepU"
/protein_id="AA04050.1"
/db_xref="GI:7263002"
/translation="MRHNFMYRLLEAGLISITSIQEGNHAKPALTRISSARSIS
PPRPVIAQOSOPAPAPITATRAELVOTLITACLDAPCRESARRTFPALTFSALDVA
RACGIAEHLTRAVFHLPAELLVHGLKSAFENIQLRYGLVACDAMGDLGFS
VATGILMAVTLPRRYLEKAGYVRLIMHDMGRNMRDINADKAGRTYNNMLHPKRS
VESEQEDVTGEGELREKSAVAVTYEVSMTAKSALPSPDMKTYRPAAPSLAENV
VWEADABETPRAYRAELVDROARALAAECGDSISCPNKLIMNTTRADACRDV
SDVGAVLVRLADVKKHOTACGTPPRNLAAYNALADLGRLEKEGQVGVDPQ
AOPAA"
complement(8717..8725)
/protein_bind
/feature
misc_feature
9164..9637
/note="region of low %G+C; possibly involved in
replication initiation"
RBS
10614..10619
/gene="resu"
10614..11590
/gene="resu"
10625..11590
/gene="resu"
/note="resolution of plasmid multimers - plasmid

```

```

Query Match      64.0%; Score 16; DB 12; Length 11910;
Best Local Similarity 100.0%; Pred. NO. 41;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 2 catgcgcagcattcg 17
|||||
DB 11110 CATGCCAGCATTTCG 11095

```

```

RESULT 28
LOCUS HS272EB/c 117968 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from clone 272EB on chromosome Xp22.13-22.31.
Contains a pseudogene similar to MDM2-like P53-binding protein
gene. Contains STSS, GSSS and a CA repeat polymorphism, complete
sequence.
ACCESSION Z93929
VERSION Z93929.1 GI:3425887
KEYWORDS HTG; CA repeat polymorphism; MDM2-like P53-binding protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 117968)
AUTHORS Grafham, D.
TITLE Direct Submission
JOURNAL Submitted (17-AUG-1998) E-mail enquiries: humquerry@sanger.ac.uk
COMMENT Clone requests: clonerequests@sanger.ac.uk
On Aug 18, 1998 this sequence version replaced gi:3334544.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
IMPORTANT: This sequence is not the entire insert of clone 272EB.
It may be shorter because we only sequence overlapping sections
neighbouring submissions.
The true left end of clone 410B11 (Z86063) is at 117869 in this
sequence. The true right end of clone 390N22 (AL008711) is at 61633
in this sequence.

```

FEATURES

```

source
1..117968
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="p22.13-22.31"
/clone="RP1-272E8"
/clone_1lb="RP1-1"
1370..1386
/note="MER20 repeat: matches 2..218 of consensus"
complement(3645..3946)
/note="MER33 repeat: matches 324..2 of consensus"
complement(3864..3994)
/note="MER3 repeat: matches 131..1 of consensus"
4194..4555
/note="MIR1A1 repeat: matches 5..365 of consensus"
5433..5531
/note="MIR repeat: matches 87..188 of consensus"
complement(6029..6111)
/note="MIR repeat: matches 188..103 of consensus"
6205..6322
/note="MIR repeat: matches 106..256 of consensus"
7754
8261..9521
/note="predicted CpG island"
complement(9741..9813)
/note="MIR2 repeat: matches 146..74 of consensus"
complement(11533..11605)
/note="MIR2 repeat: matches 145..74 of consensus"
complement(18288..19413)
/note="LIMD2 repeat: matches 1088..5 of consensus"
complement(19266..19525)
/note="L1 repeat: matches 5390..5135 of consensus"
complement(19531..22106)
/note="L1 repeat: matches 4931..2303 of consensus"
complement(22115..22419)
/note="AluX repeat: matches 301..1 of consensus"
complement(22424..22641)
/note="L1 repeat: matches 2314..2093 of consensus"
complement(23516..23641)
/note="MER25 repeat: matches 2036..1908 of consensus"
25175..25396
/note="match: GSS UB2600"
27948..28235
/note="AluY repeat: matches 1..301 of consensus"
complement(28665..29020)
/note="THB1B repeat: matches 364..1 of consensus"
complement(29889..30189)
/note="AluJo repeat: matches 302..1 of consensus"
complement(31338..31487)
/note="MER5A repeat: matches 155..4 of consensus"
33121..33328
/note="MIR repeat: matches 63..262 of consensus"
33424..33723
/note="AluX repeat: matches 2..301 of consensus"
35811..36234
/note="L1 repeat: matches 1575..2001 of consensus"
36723..39034
/note="L1 repeat: matches 2121..4429 of consensus"
37893..38188
repeat_region

```

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/MGP/ChX>

272EB is from the library RPc11 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:PCRAC2>.

```

repeat_region /note="AluX repeat: matches 1. .302 of consensus"
39040. .39339
/note="AluX repeat: matches 1. .301 of consensus"
repeat_region 39353. 40296
/note="L1 repeat: matches 4449. .5390 of consensus"
repeat_region 40148. .41111
/note="L1MD1 repeat: matches 1. .968 of consensus"
repeat_region complement(41972. .42157)
/note="MIR1 repeat: matches 541. .354 of consensus"
repeat_region complement(43606. .43671)
/note="MIR2 repeat: matches 134. .71 of consensus"
repeat_region 45479. .45585
/note="MER4 repeat: matches 26. .133 of consensus"
repeat_region 48242. .48312
/note="MER45 repeat: matches 98. .176 of consensus"
repeat_region 49604. .49826
/note="AluB repeat: matches 82. .301 of consensus"
repeat_region 49936. .50472
/note="L1PA13 repeat: matches 41. .573 of consensus"
repeat_region complement(50932. .51134)
/note="MER20 repeat: matches 218. .18 of consensus"
repeat_region 51378. .51596
/note="MER20 repeat: matches 1. .215 of consensus"
repeat_region complement(52293. .52552)
/note="MLT1C repeat: matches 466. .186 of consensus"
repeat_region complement(52555. .52908)
/note="THE1B repeat: matches 362. .1 of consensus"
repeat_region complement(52914. .53071)
/note="MLT1C repeat: matches 182. .8 of consensus"
repeat_region 53212. .53621
/note="MSTC repeat: matches 1. .403 of consensus"
repeat_region 55202. .55295
/note="MIR2 repeat: matches 1. .94 of consensus"
repeat_region 55296. .55504
/note="MER3 repeat: matches 1. .209 of consensus"
repeat_region complement(55783. .56684)
/note="L1PA10 repeat: matches 911. .1 of consensus"
repeat_region complement(56534. .57144)
/note="L1 repeat: matches 5390. .4785 of consensus"
repeat_region complement(57148. .59003)
/note="L1 repeat: matches 4720. .2850 of consensus"
unsure 59735. .59757
61714. .61945
/misc_feature /note="match: SRS 124834"
repeat_region complement(63228. .63358)
/note="MIR repeat: matches 247. .100 of consensus"
repeat_region complement(64335. .64455)
/note="L1MD3 repeat: matches 909. .790 of consensus"
repeat_region complement(65455. .65744)
/note="AluJo repeat: matches 289. .2 of consensus"
repeat_region 66343. .66557
/note="MIR repeat: matches 41. .261 of consensus"
repeat_region 66912. .67656
/note="L1MC3 repeat: matches 1660. .2409 of consensus"
repeat_region complement(68490. .68794)
/note="AluY repeat: matches 301. .1 of consensus"
repeat_region complement(68873. .69198)
/note="AluSg repeat: matches 298. .1 of consensus"
repeat_region 69217. .69532
/note="MER33 repeat: matches 1. .324 of consensus"
repeat_region complement(69638. .69703)
/note="MIR repeat: matches 129. .64 of consensus"
repeat_region 71210. .71340
/note="MIR repeat: matches 83. .206 of consensus"
misc_feature 72426. .72792
/note="match: GSS AG053969"
repeat_region complement(72923. .72973)
/note="MIR repeat: matches 139. .89 of consensus"
repeat_region complement(75007. .75196)
/note="MER5A repeat: matches 189. .1 of consensus"
repeat_region complement(75386. .75425)
/note="L1MD3 repeat: matches 406. .367 of consensus"
75508. .75800

```

```

repeat_region /note="AluJo repeat: matches 1. .288 of consensus"
complement(76744. .76840)
repeat_region /note="MLT1F repeat: matches 182. .91 of consensus"
complement(78342. .78521)
repeat_region /note="MIR repeat: matches 216. .35 of consensus"
complement(78988. .79020)
repeat_region /note="MIR2 repeat: matches 144. .112 of consensus"
complement(79137. .79186)
repeat_region /note="MIR2 repeat: matches 134. .85 of consensus"
complement(79151. .79393)
repeat_region /note="MIR repeat: matches 242. .3 of consensus"
complement(79492. .79666)
repeat_region /note="MER5B repeat: matches 178. .5 of consensus"
complement(79777. .79863)
repeat_region /note="MER5A repeat: matches 109. .27 of consensus"
79894. .79970
/note="MER5A repeat: matches 11. .91 of consensus"
repeat_region complement(80377. .80431)
/note="MIR2 repeat: matches 141. .87 of consensus"
repeat_region complement(80389. .80462)
/note="MIR repeat: matches 252. .177 of consensus"
repeat_region complement(80455. .80578)
/note="MIR repeat: matches 147. .29 of consensus"

Query Match 64.0%; Score 16; DB 9; Length 117968;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atccgcagatctcg 18
Db 97872 ATCCGCAGATCTCGA 97857

RESULT 29
AC017911 AC017911 HTG 09-DEC-1999
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
DEFINITION
ACCESSION AC017911
VERSION AC017911.1 GI:6553279
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 118100)
REFERENCE Adams,M. and Venter,J.C.
AUTHORS Direct Submission
TITLE Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
JOURNAL Rockville, MD, USA
COMMENT This sequence was identified as CDM:10212606 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source 1. 118100
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 32824 a 26059 c 26550 g 32667 t
ORIGIN
Query Match 64.0%; Score 16; DB 2; Length 118100;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 gaattcgatctcg 25
Db 63425 GCATTCGATCTCGT 63440

```

```

RESULT 30
AP002817/c
LOCUS      AP002817      140825 bp      DNA      PLN      12-AUG-2000
DEFINITION Oryza sativa genomic DNA, chromosome 1, PAC clone: p0699D11.
ACCESSION  AP002817
VERSION     AP002817.1
KEYWORDS   GI:9558510
SOURCE      Oryza sativa (cultivar: Nipponbare) DNA, clone: p0699D11.
ORGANISM   Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE   1 (bases 1 to 140825)
            Sasaki, T., Matsumoto, T. and Yamamoto, K.
            Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
            clone: p0699D11
            Published only in Database (2000) In press
REFERENCE   2 (bases 1 to 140825)
            Sasaki, T., Matsumoto, T. and Yamamoto, K.
            Direct Submission
            Submitted (26-JUL-2000) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannonda;
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail: tsasaki@abr.affrc.go.jp, URL: http://irp.dna.affrc.go.jp/,
            Tel: 81-298-38-7441, Fax: 81-298-38-7468)
            The orientation of the sequence is from SP6 to T7 of the PAC clone.
            Genes were predicted from the integrated results of the
            following: GENSCAN1.0, BLASTN2.0, BLASTX.0 as well as
            SplicePredictor (October1998 version). The genomic sequence was
            searched against the non-redundant database NRP (PIR, SWISSPROT,
            GENBEP, PDB) from MAFI DNABank and the cDNA sequence database at
            RGP. Protein similarities of the coding regions were searched
            against NRP with BLASTP2.0. ESTs represented the identified cDNA
            sequences using BLASTN2.0 with the corresponding DBJ accession no.
            and RGP clone ID.
            This sequence of p0699D11 clone has an overlap with p0462H08 clone,
            DBJ: A002525 at the 5' end. The sequence of this clone starts at
            the position 114,739 of p0462H08. The sequence of p0699D11 clone
            has an overlap with p0469E09 clone, DBJ: A001366 at the 3' end.
            The sequence of this clone ends at the position 59,396 of p0469E09.
            Detailed information on assemble quality together with annotation
            of this entry at
            http://irp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES             source
     Location/Qualifiers
         1..140825
             /organism="Oryza sativa"
             /cultivar="Nipponbare"
             /db_xref="taxon:4530"
             /chromosome="1"
             /clone="p0699D11"
             /join(527..679,925..1114,2076..2467,3067..3555)
             /note="EST C22619 (S11214) corresponds to a region of the
             predicted gene.
             Similar to Arabidopsis thaliana chromosome 2, BAC clone
             T1D12, putative protein kinase (AC006587)."
             /codon_start=1
             /protein_id="BAB03429.1"
             /db_xref="GI:9558511"
             /translation="MWGIRRNKGHEATFTLLELATDNNTSTECILGRGSGSYK
             APLNDREYVAVKOLDLNGLOGNEPTEFLVEYLMISLHHPNLYKFGCVGDDORLLYE
             YMLGSLDEDLRLRGOELRDMETRMKTAADAAGALEYLLHDEATPAVLYRDRPSNI
             LLEGYNKSLGKLGKLGVDGKTHVTRMGTHGCAPEYLSGKLITKSDISFG
             VVLELITGRRLDSNRPRDEDLVAMARPLFKDGRPKMADPFLSHGFPRKGFQA
             LAIAAMCLOEKAKNRPISREYVALSYLASQTHESQNTAARHTLPQSPRYLDNDIN
             ODTSLRSGVHMPPLAGTIDHMQVEVKEMCRSSHPRGRTVPGADRRRLAENAV
             WEAMRERKSTSMR"
             /complement(8769..9326)
             /note="ESTs AU093260(C52190), AU062991(C52190) correspond
             to a region of the predicted gene.
             hypothetical protein"
             /codon_start=1
             /protein_id="BAB03430.1"
CDS
     1..140825
         /translation="MDDPLEPTLELCRAFTIIRGTSQKIVALSKEPALLVDSAFMSI
         AVDOQSCVMDLDFGNGMGIFLIDVGGSDHPIIGISDGLVCLLKAPRAPIVGLLTN
         LPRKLSVLDLAKHGCTLRRLALYCLAFSCROYLAQGMIVSISLSALMREVALVSFK
         GSHGCLADAASYLGAELORLPRCG"
         /complement(join(15587..15727,15868..15915))
         /note="hypothetical protein"
         /codon_start=1
         /protein_id="BAB03431.1"
         /db_xref="GI:9558513"
         /translation="MNNFHYEALIGRGNSTVYKGRKKSIIEYFAVKSVDKRSKVL
         NEVSAAPSPCSFSSW"
         /complement(join(18733..18801,19343..19489,19843..19905,
         20016..20087,20611..20691,20769..20819,20976..21072,
         21611..21753))
         /note="ESTs C72327(E1437), AU078752(E1437) correspond to a
         region of the predicted gene.
         Similar to Archaeoglobus fulgidus competence-damage
         protein (AE000949)"
         /codon_start=1
         /protein_id="BAB03432.1"
         /db_xref="GI:9558514"
         /translation="MLSDGRLERAGRTKKNISSVSGTSTNEVEQTI SRSAIIIVG
         DETLFGVEDEKLGAGLCKRIHLAIGRWVSHVAVVSNIDVAEVEERCKSTDMDVFLVG
         GLAPLHSDISLACVAKAFGVRLAPDEFEFEYSQLDGNYTDRDNEMALLPGITELL
         HHMLPLPLICKKNVYLIATNVDELTEWGLDTPQESGLVMKSFYSKHLCTSLD
         VKIAPYAKICIDFSDVYIGMT"
         /join(30551..30609,30919..31139,31223..31578)
         /note="Similar to Arabidopsis thaliana chromosome 4, BAC
         clone F10M23; hypothetical protein (AL035440)"
         /codon_start=1
         /protein_id="BAB03433.1"
         /db_xref="GI:9558515"
         /translation="MLGKGRKRRREGGENDIYAKKRSGMDPLPHLRGKSTSKVP
         OLVTGILKNVLEPQNRVAGSLNCTKTRTRSSVYLPVLYTVAESEMRRHVKRGGH
         HDLIRQGGAGAEERDEGVYTDGGLQHRQPHRYIPSEAAVYRRSGMRATOSVGGED
         GDAHAYSHKAVARRAANSHSVLEAGAGMLKGCGLHHVHNVAVLCRTGFLD"
         /complement(join(31912..32455,32698..33684,34091..34341))
         /note="hypothetical protein"
         /codon_start=1
         /protein_id="BAB03434.1"
         /db_xref="GI:9558516"
         /translation="WRKKRTNLLIVGDEATFFFEERCRRRAGVSGSEVERBDSKP
         SRGRHPPATRAAPLRKQKARDEDLATSSGTPSPGMRPHHPLPSSPFLSLSH
         SLLIFSGRIGEGHCEAASACSEGVASGCGAATAAAGACRGCGCAPPTSP
         RHPALSSAAPGARDADAGLVGCGEGGTGAGARASRRRQIGQENGGHGRAPYS
         VCCPFTYVLAELVLAIVCVPAALCHLAVRGHRRRVGSAKOKEMGELLTIDYSPS
         LAAAKARKVEVEFPATPMAEHIGEAFAWMKSCLOALHMYARLLSSAGEDARPA
         DERVLPLVACPLSPVPTLPRHPRVRSLSCLPRKLASTLSTPQRRGRNDSRKHP
         TEGMKIKERRRNSITIAQKTRSRKAPRTMPVVARASATVPNDSDRWGAGAPARSAR
         BEGGGGGGDADALPACRLILAAVSLAASPPKYTEPKPVRVPLTNNHPIRGAAPPCR
         SSSPFLSSSSLSRSGSGKPARCGVYARERLRLWRVROLKATSIHAERAAHVIVVVLV
         VKLPAREACPRSTVPASRGPPHRR"
         /join(35382..35472,36089..36172,36426..36530,37078..37337,
         38995..39114)
         /note="hypothetical protein"
         /codon_start=1
         /protein_id="BAB03435.1"
         /db_xref="GI:9558517"
         /translation="MQLDASDFHGOELTLKLRHDKLSLESSGKSYDLVSPFAOP
         DATVFLSPGPAKAVGKIARVSLVRYRDEPELEKPLGSLTSPSSKASSSKTKR
         FTSGSKNSSGCSAOSISIGOOSAEFAHKNKRDDESLGINSVSGSABESQVAGGDS
         GTSEVQOTVPEKSKKKKKKIDRPRWKRRTPTDRKKTQOELDRVAVDSSLYKSPPKY
         O"
         /join(40731..40753,43157..43538,44516..45821,47347..47893,
         48657..48934,48983..50395)
         /note="Similar to Arabidopsis thaliana chromosome II BAC
         clone T27A16; hypothetical protein (AC005496)"
         /codon_start=1
         /protein_id="BAB03436.1"
         /db_xref="GI:9558518"
         /translation="MYSLRPKIRPPFALAAARDSPLRPPPPGRPLPSMIRWGRD
         RRLPYAMSRGPEPGSGRCGRNDEEERSRRRGREGGQDGEDELKGRKGEKEKRE

```


EIRLACGMRBRDCTGSEFADFLSLFNKCLVSLMLTAKDEIKQLAGNRPRDGTGIC
 HGAACKDLPIIDRSIYEAKIAGHSIHOLASCWMLHJGSCCKAGLMRRSKSASAL
 LISCSPWNPTEPCPNRSOPATTLATSIISFPPSYLYFSLONYISCFIISI
 FSLNLFGRSPSPPTSIVRPPTAMTAVRSIVIPYDGVAREMKLAPTSIISMS
 SLINPMEOYRCQSSALSISLGTPTIHRBAKPIAFEGCNPFPEFLKTKSTALAH
 AYTHSSSINKRATPTIRKIDILLYKORAPULCTSPGNSRHHIINSHOCHTHRLPJO
 YRVLSSEHARKOREAREBTAPPHRTSISEPNRPPATPAREMPPPPPPSLP
 DLHRAHAYLIPRGHRAHARASALLASLRPLDHRALAVRVHVSIRAAARLPG
 RGGGLAALHSLILVRAGHADPHASASILVQVSCSVASARVDEETASADVSM
 NMVIDYVKSGDLARARELFDVMGRNVSVTWVIGAYAQMKOPEEAIEVRNMOEG
 IEPDVALISVACGDLGVYDGLGEMVFRVLRGLCOEIPLMMAIIIDVFKCSYVK
 ALFEPCEMOKSYVTMTMAGFALHVGLEAVEFPRMEKENSMDITFLAVSYC
 SHVGLIDGRNFKFTMVSQYKIKRVVHVGCMIDILGRACLMARAGLIDMPKANA
 AIWGLALLAARTHGTDELGBOLLLHLELPHNSGNTILLSNIYAEQERDVAPELPI
 SMRGRLENNVPGASSIDVGMVHEFTSRDGSHPSLHIREVLCATINSIKSVGHIALI
 PESLHDVEEG"

CDS
 complement(join(52441..53547,53887..54014,54128..54165,
 54272..54487,54709..54870,55838..56016,56150..56256,
 56375..56539,56658..56741,56821..57040))
 /note="Similar to Arabidopsis thaliana chromosome I BAC
 clone T5A14; Kinesin (AC005223)"
 /codon_start=1
 /protein_id="BAB03437.1"
 /db_xref="GI:9558519"
 /translation="MDVQPARTMRLPDTLSLMGFNKLTPPSIESVSHIIDLSPT
 KROMVYKEDENISDDMTESAKYOKTODELSVLAQLKITTORREALNNYDLKG
 NTRVFCRIKPEHIESISSRLFTIDESNVELKAVETKRYOKKIDKVDQSTGQDV
 SEVPVKSALDGVVVCIFAYGQTSQKTYMEGKPENLGVIPRGITOLFPOASCCN
 RELFTFPMLEIYMGNIIDRLAPRSKNGIKNPSLISKSDPGGIEIDLVAVNS
 OEVRLYEMGTLRSTASTMANSTSRSHCLIRISLISLNTERKATSKLMIIDLGG
 SFRILKVTATGRLEKGAINTLSALGVIDALQTKRPHPYRNSKLTQVLOSLGQ
 ESKLTAVHISPDGDLCTICTGLFATRVRSIRLESEPPMKARKETLLDLOGV
 NDLEHCEDEIRIKINLESMEHLTPOPTIYSNFMDSISEELKTIVSVSVYKNA
 REASRRDPRMKPTASSOHRIGLNKRPFIINRLKPPVPRRRSSVAYSSVYNA
 AEWQSCSESCSLSDMNWTPSLRDGECSDASETEIKOYIFSEHKSHDYVC
 YTDPLAESRDIOIKIEKGIVDIDNMUHQIIVERTSIFRSKMLVDIFGVETAEIHS
 SIPTPTTACTKEDSQVDEVMGLTLOSSTDYVEDIKOSTDNOFTAKELTPPKPEF
 SSNNEVKGHKNEHPYHGRPRRSIQELENCLEKPNMDSKSHRSHDKHKTGNVLSY
 TGE"

CDS
 join(57603..57645,57742..57843,58195..58344,59425..59479,
 59537..59589,61304..61785)
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAB03438.1"
 /db_xref="GI:9558520"
 /translation="MKQDKGNRRCLPLOGSFSAPIYHOVHTIKGECCPTDPSEBSR
 SETAKLMDIDIPWMAITTPNGMKRCLANENHSANITRKGSPTMCHALEFSEATAIS
 LIAFLAGKAPFELSKRRSGDVHMAISGVEAAAPLPPOPPILPPGDKMPAAG
 LCLSPASARLRSPLPHQRQPPMRDRAAPAAALRLSPAHRRRETIKRLPLSTSP
 AIORPCMRPLPADATGELTELIVSSKCRSCVEFIPSTVELLPMWGELELFSSTVKL
 LPSPCSAPMSLAPSSRA"

CDS
 complement(62330..63121)
 /note="Similar to Arabidopsis thaliana DNA chromosome 4,
 BAC clone M3E9; hypothetical protein (Al022223)"
 /codon_start=1
 /protein_id="BAB03439.1"
 /db_xref="GI:9558521"
 /db_xref="GI:9558521"

Query Match
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 Oy 2 catgccagcatccg 17
 |||||||||||||
 Db 89622 CATGCCAGCATTTGC 89607

RESULT 31
 AP001366/c
 LOCUS AP001366 146081 bp DNA PLN 12-AUG-2000
 DEFINITION Oryza sativa genomic DNA, chromosome 1, PAC clone: P0469E09.
 ACCESSION AP001366
 VERSION AP001366.1 GI:7228436

KEYWORDS
 SOURCE
 ORGANISM
 Oryza sativa (cultivar: Nipponbare) DNA, clone: P0469E09.
 Oryza sativa
 Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
 Eriophytidae; Oryzoideae; Oryza.
 1 (bases 1 to 146081)
 Sasaki, T., Matsumoto, T. and Yamamoto, K.
 Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 clone: P0469E09
 Published only in Database (2000) In press
 2 (bases 1 to 146081)
 Sasaki, T., Matsumoto, T. and Yamamoto, K.
 Direct Submission
 Submitted (09-MAR-2000) to the DDBJ/EMBL/Genbank databases. Takuji
 Sasaki, National Institute of Agrobiological Resources, Rice Genome
 Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail: tsasaki@abrc.affrc.go.jp, url: http://rpg.dna.affrc.go.jp/
 Tel: 81-298-38-7441, Fax: 81-298-38-7468)
 The orientation of the sequence is from SP6 to T7 of the PAC clone.
 Genes were predicted from the integrated results of the
 following: GENSCAN 1.0, BLASTN 2.0, BLASTX 2.0 as well as
 SplicePredictor (October 1998 version). The genomic sequence was
 searched against the non-redundant database NRP (PIR, SWISSPROT,
 GENPEPT, PDB) from MARF DNA bank and the cDNA sequence database at
 RGP. Protein similarities of the coding regions were searched
 against NRP with BLASTP 2.0. ESTs represent the identified cDNA
 sequences using BLASTN 2.0 with the corresponding DDBJ accession
 no. and RGP clone ID.
 Detailed information on overlap and assemble quality together with
 annotation of this entry at
 http://www.dna.affrc.go.jp:82/genomicdata/genomefinished.html.

FEATURES
 source
 location/Qualifiers
 1..146081
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /chromosome="1"
 /clone="P0469E09"
 join(5902..5980,6229..6272,8134..8295)
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAA92397.1"
 /db_xref="GI:7228437"
 /translation="MGAYKRSSFLVVVVFALLITSMAGGRKMLINKHOVSNETS
 /DESMHOGQEDDEMAMVHERILKOVKTNDYDPTPTAKPIAKETLPN"
 9673..9989
 /note="5' LTR"
 join(10528..12588,12673..13083,13175..13473,13550..13808,
 13901..14083,14241..14306,14743..15045,15216..15479)
 /note="EST C28952(C62945) corresponds to a region of the
 predicted gene.
 Similar to maize transposon MUDR mudra protein isolate
 (AC003981)"
 /codon_start=1
 /protein_id="BAA92398.1"
 /db_xref="GI:7228438"
 /translation="WTSKGIIDRPARTPOSIIYSHLIRFRITDOEYTWMSVVSRAI
 EGYFWELMPMDSTAAMRVRVEMAFERSPVLTVEVSOKRDINVSQOTDVGCPINAG
 VGVSPNOMENQPREQAMGMADEBERGIIYDEVEDERSQNEADDASSDEEDVMA
 ATDMANDEPESGLVIEGDHVPWEYKENYIDGARVAHDEKAEVAKHVAISLOREFV
 VKSTNVYVEVRCMKEDCPMRYAHYKGNNDYKWSIYIEHKCYLOGVYKRNITSAF
 VASEMSVSVNGIGEPKSIITHIENKRYTISYKACRAKOKITIEYGFEESSYN
 LPRLATTIAORNNNTYIDLTFTSYDDTKSYLOKAFSICACIAPFVCHRPVLCIDG
 TFMTRGTRCOIIITAIIGCGNNOVPMARFVSESTESMYNPLEVNIYAVYMRNVC
 LIHDHAGMIRAIIDYLONGWDEKRLPAKMPDVRSWCMRHGANGFYKPKKHLMELF
 KRICAKNCEKFNELMDKLDLITTKOTDEOSRRPQVEDDEPILPGALHDPPTMRK
 SCSSTINFTOWIENETPEKWSLLPDTDSRGYIMTINLAEVYNNMVRGVRLPLVAIV
 EPIHRTQVAFEDRYKTKIGPSMADNNIYGEFQVWRPLPHLGPANTLPPYSTVGSKAN
 LIQFTADILHDPADNNNSRSLSLXATLLMLFCWMTSTSHGAHVDRLHYARISIVAO
 PDVDPWMSGSAVLAATYRALCEACTKIDAGATIGCCPMLDQMAAEKPAICRPVYDS
 AYGVSRSQWPEDEGPTMGTYWCRGRGRRRAHQVRRGYPDVFEDRLQPSDYIETP
 TEBAVAVARAPILGISICTRDQAYWLTLLPMWVDIIVEPHCPQVRVROEGCLROVFPNGV

LTR	QPTVLPADHSLETRRQGLAGALNAPRVQOYVDDMLATEEVLNELLPHTEENRYDIYLM YLPRTARVATFPDAPEPHVAATDAVPPHBRDVEVAGADARDISADITTAQVRINLH GHLTVEQRTEDRMQEKRAVYRSCRSADIVPPAGPDLVCLRLPLAEECDIU HRETTVTRPCNFIEMVDMENQNDGTAYSPESRSDLRKDECECTAABALMHI VNEGLPTMRPREGCDRCVIRSVRRRTGQRCVCPNIVDDPFTIHAMTEYER RVIALSLDRNRNPLGLPKMSERRKCCCRDRCMOVAIVSAFOTFRRRYFVCPVNDNFV CLUTYKTRIDRLHITTYOTE 16155..16469	LTR	complement(43999..44385) /note="5' LTR" /join(44572..44723,44846..44938,46366..46378) /note="hypoetical protein" /codon_start=1 /protein_id="BAA92403.1" /db_xref="GI:7228443"				
CDS	/note="3' LTR" /join(23418..23748,25558..25589,25715..26480,26838..26965) /note="hypoetical protein" /codon_start=1 /protein_id="BAA92399.1" /db_xref="GI:7228439"	CDS	/translation="MALPKGVRSFKPMAAASASPPQPGHMLVIGTGVGRVYSQ RLAOGVRSQTCNSPAKTELEMLGADASIFPATSSRCVW" complement(48119..48257,48429..48703,49975..50139) /note="hypoetical protein" /codon_start=1 /protein_id="BAA92404.1" /db_xref="GI:7228444"				
LTR	/translation="MGANCIATKAKERPQCVPIEVSARVNRVHSPSMFRNDRTIH EDIMEMPALFESSHSGSIRPETSGSIAPTDGSNGSPDMFNKLCCHKSPRRESS KIASLDLRSTTSSNSPEAKLSRKSLDVSASDKMSITSPPTPAITSNADSSSS RHSLSLPTDADSMKARSPGQLYROVSDSKIPSLRSLNEGSPGRTSSMSLCSN DLAAGSHGESSDWSMRTPEMVAVSSQBRNSVDELGSVSSKMTSSNANPCTH SDPOVCKLCKLKERSTWAOELAVVAALCGHVHADCLDLSLAEDRYDDPCPV CTHGEKCTVCKLGLSKSTKNKIPKNVIVDVALDGSXKHOKEKIPLEBELYSQARASH RELKSSRIGANDMTWLLNDLRLEGS" complement(31926..31991) /note="ESTs C96615(C10106), C26336(C12127), D21959(C10106) correspond to a region of the predicted gene. Similar to NAM (A0021889)" /codon_start=1 /protein_id="BAA92400.1" /db_xref="GI:7228440"	CDS	/translation="MALVINKDESDVDVNAFACARRAPLVTCGPMGFVOLARRLSEC EPDFDRKAAIIISFCALSLFESKNTRIYIVLLCYLRKNRPLIVTESKRRKCRIT ASLASRIGDELVCRLARLASALDRVFKILKTSDFVLMVLRKCIIVNPVHISRTH GSSMYSTLPVAVYRSTERRLAGIVIDCTIG" complement(51918..52637,52684..53346) /note="Similar to Arabidopsis thaliana chromosome 1 YAC YU8BH12R sequence, C2-HC type zinc finger protein C.e-Wy11. (AC002986)" /codon_start=1 /protein_id="BAA92405.1" /db_xref="GI:7228445"				
CDS	/translation="MESLUDMLPPGFGFHPKDTLELSHYLKKIHGOKIEXETIPEV DIKHEPDLPAKCDVPTDQNMHFFARDKRYPNGSKSNATAGWAKSTGKRAIK MGKQIGTKTFLVEHGRPPGRTTMEHYI DERCOACPRMKATVLCRTIKTN DWIPNGEINLSDPPEPPYDAPSPYSTEDLNPADEVGVGAAPVAVPEQVTT AITANIPSPSDINDMLNELPFPDPOSLASADLSPDEQVSSNGALAKVE ODSPSPNVDVDTIELDIPVNI LHPGDDFMNLQNPDDOPIQVATVWSGLOKE ELMSFOAARPSQSNADNGITIRRYRSMKTPETSPQFKQAKMRGINKKATSS SESJNCTIKFENSGRLVHOHCAHDVAATSTPQKQYDQAKMRGINKKATSS CSDAKMWILVAGFAIGAVVAALHIGQRLGSLGSDQH" complement(38747..38998) /note="hypoetical protein" /codon_start=1 /protein_id="BAA92401.1" /db_xref="GI:7228441"	CDS	/translation="MAGHPRYLADTARPDMMWAVGPEVRHMDRIGTPPEVYILCRHD DRGRASILAMMRMLGLELVHFGCCYDDEPHGHSCTS" complement(39315..39726) /note="3' LTR" complement(40209..40396,40917..43545) /note="EST C28952(C62945) corresponds to a region of the predicted gene. Similar to maize transposon MUDR mudra protein isolog. (AC003981)" /codon_start=1 /protein_id="BAA92402.1" /db_xref="GI:7228442"	CDS	/translation="MAGHPRYLADTARPDMMWAVGPEVRHMDRIGTPPEVYILCRHD DRGRASILAMMRMLGLELVHFGCCYDDEPHGHSCTS" complement(39315..39726) /note="3' LTR" complement(40209..40396,40917..43545) /note="EST C28952(C62945) corresponds to a region of the predicted gene. Similar to maize transposon MUDR mudra protein isolog. (AC003981)" /codon_start=1 /protein_id="BAA92402.1" /db_xref="GI:7228442"		
LTR	/translation="MSSKATFOIVHSEGNIRESPGVDSDFVMTSGIDRPAERTFO SITSMILGRFRIDQEVYITMSVAVVSRATEGYFWEMLPMDSTDMRVEYEMFEMWMP LIVFVSVOEKDINVSQTEDESPINAGDVVGPSSQMNENQPREEQAMQAMGDEPVC IIVDEMEREDSDNEVDDASDEGDVATDMANDESGILVISEGDHNPVEMKREY IEGARVAHDKEMKAVKHAVALSOREFRVAVKSYVYEVRCMKEDCPMRVHAKWYN DYKVSIVTEHCKCYLOGVEKHNITISAMASMYSVVGNIGFEKSTIRHINENKFN YITSAKAMARAKOITILEMRGFEASDYNDLRLAIIAORNNITYIDLHTFTSVDRIT KYLORAFESTICINAFVHCPVYLCIDTETFGTKRGQILVATGNGNNOVLPAALIA FVSESENTESYFLEHVAIVVMRPVNCILIDHRAAGMLRAIDYLQNGCNDGLPAALIA PDVRSFMCIMHGANFYKQKKNHLMELFKLCAQOEFKFNELMKLDELITTKQDKE OSRRPLVEGDEDEPILGALHDPPTMRSSSSIRNFQWIMENREKKSLLPDDDS RGVIMTNTLAEVNVMRGVRYLPLVAIVEFLHGOAVFERNREKKSLLPDDDS GNVYTKYMEKIKKARRHVAOGTQVHYELMCDRBSRGITRKQAVDEGLADGG CTCSMKPKLIHHPCSHVLAAGDCGIPPNVIVSNFRKALFTHWSEIYFGTSGS YITLSAOVYITDPKSLRVKGRGRTIRLNMDSSEAGRLKCSKCDLRGHTYKIC PKRAEVPSCGADSPGQSDVLFJHMYLEQMLTTHFRCTVVSDDPAKRVHNSITHH KQLELSTFGAAYLAKVNAHA"	LTR	complement(57570..58009) /note="3' LTR" complement(58018..58891,58980..62098) /note="Similar to Arabidopsis thaliana chromosome II BAC F9013 genomic sequence, putative retroelement pol polyprotein. (AC006248)" /codon_start=1 /protein_id="BAA92406.1" /db_xref="GI:7228446"	LTR	/translation="WRVNLQAGCLMTAIDPGVAFERRDWAALSLIQAVPREMLRGLA KHDTAAAMDALIKTRVGVDRVREKKEGDFRQFSMFKREPEERFAMLTAVVAD IDMGVMEDEHVNKRLRYVKKYKPAISLEOLDVYTMALTEIRKSLSTVSYSD IDEGSDGKLYLITEOMQARVYKORQEGSSGNGKGRAPCTONHKKRGPGSGKRA ATGAUSSRDISKVCNCFNCDFGHYAROCRRPRORGRANIVQAADEEPTLLMAHVG VSLAGEATLIGRPGQOEYHLEKRYLIDHDEGCEYEYTRDFLDTGATNHTGVRSAP AELDTCVGVGYKFGISVIELQGRSTVVFRCNCDHNSLDIVYIIPKLRNIIISGRIL complement(57570..58009) /note="3' LTR" complement(58018..58891,58980..62098) /note="Similar to Arabidopsis thaliana chromosome II BAC F9013 genomic sequence, putative retroelement pol polyprotein. (AC006248)" /codon_start=1 /protein_id="BAA92406.1" /db_xref="GI:7228446"		
CDS	/translation="MAGHPRYLADTARPDMMWAVGPEVRHMDRIGTPPEVYILCRHD DRGRASILAMMRMLGLELVHFGCCYDDEPHGHSCTS" complement(39315..39726) /note="3' LTR" complement(40209..40396,40917..43545) /note="EST C28952(C62945) corresponds to a region of the predicted gene. Similar to maize transposon MUDR mudra protein isolog. (AC003981)" /codon_start=1 /protein_id="BAA92402.1" /db_xref="GI:7228442"	CDS	/translation="MAGHPRYLADTARPDMMWAVGPEVRHMDRIGTPPEVYILCRHD DRGRASILAMMRMLGLELVHFGCCYDDEPHGHSCTS" complement(39315..39726) /note="3' LTR" complement(40209..40396,40917..43545) /note="EST C28952(C62945) corresponds to a region of the predicted gene. Similar to maize transposon MUDR mudra protein isolog. (AC003981)" /codon_start=1 /protein_id="BAA92402.1" /db_xref="GI:7228442"	CDS	/translation="MAGHPRYLADTARPDMMWAVGPEVRHMDRIGTPPEVYILCRHD DRGRASILAMMRMLGLELVHFGCCYDDEPHGHSCTS" complement(39315..39726) /note="3' LTR" complement(40209..40396,40917..43545) /note="EST C28952(C62945) corresponds to a region of the predicted gene. Similar to maize transposon MUDR mudra protein isolog. (AC003981)" /codon_start=1 /protein_id="BAA92402.1" /db_xref="GI:7228442"	CDS	/translation="MAGHPRYLADTARPDMMWAVGPEVRHMDRIGTPPEVYILCRHD DRGRASILAMMRMLGLELVHFGCCYDDEPHGHSCTS" complement(39315..39726) /note="3' LTR" complement(40209..40396,40917..43545) /note="EST C28952(C62945) corresponds to a region of the predicted gene. Similar to maize transposon MUDR mudra protein isolog. (AC003981)" /codon_start=1 /protein_id="BAA92402.1" /db_xref="GI:7228442"

Query Match 64.0%; Score 16; DB 8; Length 146081;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 8193 CATCGCAGACATTTCG 8178
RESULT 32
AC007476/c
LOCUS AC007476.5
DEFINITION BACR37906, complete sequence.
ACCESSION AC007476
VERSION AC007476.5
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Insecta; Diptera; Brachyera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS
1 (bases 1 to 171705)
Celisner, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanalides, P.G., Brandon, R.C.,
Rogers, Y., An, H., Baldwin, D., Banazon, J., Beeson, K.Y., Busan, D.A.,
Carlson, J.W., Center, A., Chapple, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Dou, L.F., Doyle, C., Dretnick, D., Farfan, D.,
Ferreira, S., Frisoe, E., Galle, R.F., Garg, N.S., George, R.A.,
Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
Idegawa, C., Jallali, M., Kruse, D., Li, P., Mattei, B., Moshireli, A.,
Mintosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nuno, J.,
Paclet, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,
Phonemavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
Stapleton, M., Strong, R., Swirski, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
Sequencing of Drosophila chromosome 2R, region 49A-49B
Unpublished
2 (bases 1 to 171705)
Celisner, S.E., Abdayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,
Butenhardt, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Kaira, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshireli, A.R., Moshireli, M., Nixon, K., Paclet, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequiera, A., Sethi, H., Smit, E.,
Swirski, R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
Direct Submission
Submitted (05-MAY-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Mar 17, 2001 this sequence version replaced gi:5670614.
COMMENT
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to bdgpf@fruitfly.berkeley.edu.
FEATURES
source
Location/Qualifiers
1. 171705
/organism="Drosophila melanogaster"
/strain="Y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="2R"
/map="49A-49B"
/clone="BACR37P06 (D594)"
/clone.lib="RP2-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBAC63.6)"
BASE COUNT 47654 a 38437 c 38157 g 47457 t
ORIGIN

Query Match 64.0% Score 16: DB 3: Length 171705:
Best Local Similarity 100.0%: Pred. No. 40:
Matches 16: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
QY 10 gcaatcgaatcgcgt 25
|||||
Db 106124 GCATTTCATCTCTCT 106109

RESULT 33
AC037435 183937 bp DNA HTG 15-JAN-2001
LOCUS
DEFINITION Mus musculus chromosome 11 clone RP23-9813 map 11, WORKING DRAFT
SEQUENCE, 24 unordered pieces.
AC037435 AC037435.2 GI:12229535
VERSION
KEYWORDS HTG: HTGS_PHASE1: HTGS_DRAFT.
SOURCE house mouse.

ORGANISM
Mus musculus
REFERENCE
AUTHORS
1 (bases 1 to 183937)
Birren, B., Linton, L., Nusbaum, C., and Lander, E.
Mus musculus chromosome 11, clone RP23-9813
Unpublished
2 (bases 1 to 183937)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F.,
Boguslavsky, L., Bouknight, B., Brown, A., Burkett, G.,
Campiano, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S.,
Collymore, A., Cooke, P., DeArrelano, K., Dewar, K., Diaz, J.S.,
Dodgson, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Gadgil, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grandpierre, N., Grant, G., Hages, B., Heath, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lebecky, J.,
Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Meneus, L., Milhova, T., Miranda, C., Mienga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (08-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 15, 2001 this sequence version replaced gi:7528163.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www.seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: 98_L13
Center clone name: L8882
Summary Statistics
Sequencing vector: M13: M77815: 4% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap: version 0.960731
Consensus quality: 170056 bases at least Q40
Consensus quality: 177644 bases at least Q30
Consensus quality: 180041 bases at least Q20
Insert size: 188000: agarose-fp
Insert size: 181637: sum-of-contigs
Quality coverage: 4.7 in Q20 bases; agarose-fp
Quality coverage: 4.9 in Q20 bases.
NOTE: This is a 'working draft' sequence. It currently
consists of 24 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 23881: contig of 23881 bp in length
* 23882 23981: gap of 100 bp
* 23982 24465: contig of 488 bp in length
* 24470 24569: gap of 100 bp
* 24570 24755: contig of 186 bp in length
* 24756 24855: gap of 100 bp
* 24856 25997: contig of 1142 bp in length
* 25998 26097: gap of 100 bp
* 26098 26764: contig of 667 bp in length

```

* 26765 26864: gap of 100 bp
* 26865 28426: contig of 1562 bp in length
* 28427 28526: gap of 100 bp
* 28527 30488: contig of 1962 bp in length
* 30489 30588: gap of 100 bp
* 30589 32514: contig of 1926 bp in length
* 32515 32614: gap of 100 bp
* 32615 36175: contig of 3561 bp in length
* 36176 36275: gap of 100 bp
* 36276 38734: contig of 2459 bp in length
* 38735 38834: gap of 100 bp
* 38835 40927: contig of 2093 bp in length
* 40928 41027: gap of 100 bp
* 41028 43453: contig of 2426 bp in length
* 43454 43553: gap of 100 bp
* 43554 46744: contig of 3191 bp in length
* 46745 46844: gap of 100 bp
* 46845 50369: contig of 3525 bp in length
* 50370 50469: gap of 100 bp
* 50470 56799: contig of 6330 bp in length
* 56800 56899: gap of 100 bp
* 56900 57617: contig of 718 bp in length
* 57618 57717: gap of 100 bp
* 57718 68517: contig of 10800 bp in length
* 68518 68617: gap of 100 bp
* 68618 82793: contig of 14176 bp in length
* 82794 82893: gap of 100 bp
* 82894 96371: contig of 13478 bp in length
* 96372 96471: gap of 100 bp
* 96472 112208: contig of 15737 bp in length
* 112209 112308: gap of 100 bp
* 112309 127004: contig of 14696 bp in length
* 127005 127104: gap of 100 bp
* 127105 147474: contig of 20370 bp in length
* 147475 147574: gap of 100 bp
* 147575 174126: contig of 26552 bp in length
* 174127 174226: gap of 100 bp
* 174227 183937: contig of 9711 bp in length.

```

FEATURES

```

source
1. 183937
   /organism="Mus musculus"
   /db_xref="taxon:10090"
   /chromosome="11"
   /map="11"
   /clone="RP23-9813"
   /clone_lib="RPC1-23 Female Mouse BAC"
   1. 23881
   /note="assembly_fragment"
   clone_end:SP6
   vector_side:left"
misc_feature
23982..24469
   /note="assembly_fragment"
24570..24755
   /note="assembly_fragment"
24856..25997
   /note="assembly_fragment"
26098..26764
   /note="assembly_fragment"
26865..28426
   /note="assembly_fragment"
28527..30488
   /note="assembly_fragment"
30589..32514
   /note="assembly_fragment"
32615..36175
   /note="assembly_fragment"
36276..38734
   /note="assembly_fragment"
38835..40927
   /note="assembly_fragment"
41028..43453
   /note="assembly_fragment"
43554..46744
   /note="assembly_fragment"

```

```

misc_feature /note="assembly_fragment"
46845..50369
   /note="assembly_fragment"
50470..56799
   /note="assembly_fragment"
56900..57617
   /note="assembly_fragment"
57718..68517
   /note="assembly_fragment"
68618..82793
   /note="assembly_fragment"
82894..96371
   /note="assembly_fragment"
96472..112208
   /note="assembly_fragment"
112309..127004
   /note="assembly_fragment"
127105..147474
   /note="assembly_fragment"
147575..174126
   /note="assembly_fragment"
174227..183937
   /note="assembly_fragment"
clone_end:T7
vector_side:right"
BASE COUNT 47977 a 43626 c 42743 g 47258 t 2333 others
ORIGIN

```

```

Query Match 64.0%; Score 16; DB 2: Length 183937;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcatgccagcatc 16
Db 90966 GCATGCCAGCATTTTC 90981
|||||

```

```

RESULT 34
AE003822/c DNA INV 04-OCT-2000
LOCUS
DEFINITION
AE003822 AE002787
VERSION
KEYWORDS
SOURCE
ORGANISM

```

```

REFERENCE
AUTHORS
1 (bases 1 to 266133)
Adams,M.D., Celisner,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanalides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,J.H., Blazee,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abill,J.F., Agbayani,A., An,H.J.,
Andrews-Plankkoc,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brotlier,P., Burris,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
Fleischmann,W., Foster,C., Gabrielian,A.E., Gary,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostlin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,

```

Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Liu, X., Liu, X., Matel, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Pollard, J., Puri, V., Reese, M.G., Rehnert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svitskas, R., Tector, C., Turner, R., Venier, E., Wang, A.H., Wang, X., Wang, Z., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yen, R.F., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)

2 (bases 1 to 266133)
20196006

JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
SOURCE

Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, J.C.
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7303437.
Location/Qualifiers
1. 266133
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2R"
join(392..425,1007..1155,5710..6105)
/gene="Cam"
/product="CT41657"
/db_xref="FLYBASE:FBan0008472"
/db_xref="FLYBASE:FBgn0000253"
<392..>15192
/gene="Cam"
/note="CG8472"
/map="48F5-48F5"
/db_xref="FLYBASE:FBan0008472"
/db_xref="FLYBASE:FBgn0000253"
join(405..425,1007..1155,5710..5884,10160..10402,14732..15192)
/gene="Cam"
/product="CT41655"
/db_xref="FLYBASE:FBan0008472"
/db_xref="FLYBASE:FBgn0000253"
join(596..619,1004..1155,5710..5884,10160..10402,14732..14876)
/gene="Cam"
/product="CT41659"
/db_xref="FLYBASE:FBan0008472"
/db_xref="FLYBASE:FBgn0000253"
join(1153..1155,5710..5922)
/gene="Cam"
/note="Cam gene product [alt 1]"
/codon_start=1
/db_xref="FLYBASE:FBan0008472"
/db_xref="FLYBASE:FBgn0000253"
protein_id="AAF58544.1"
/db_xref="GI:7303488"
/translation="MADOLTEEOIAEFKFAFLFDKDGDTITTKELGTVMRSLGONPTEAFLODMINEVDADNGTIDPEFLTMARKKDDTSEEIRARFVPDKDNGFISAAELRHVMTNGEKLTDPEVDMEIRADIDGQVNEEFVTMTSK"
/gene="Cam"
/product="CT41657"
/db_xref="FLYBASE:FBan0008472"
/db_xref="FLYBASE:FBgn0000253"
join(5527..5642,5710..5884,10160..10402,14732..15192)
/gene="Cam"
/product="CT41655"
/db_xref="FLYBASE:FBan0008472"
/db_xref="FLYBASE:FBgn0000253"
join(5640..5642,5710..5884,10160..10402,14732..14760)
/gene="Cam"
/note="Cam gene product [alt 3]"
/codon_start=1
/db_xref="FLYBASE:FBan0008472"
/db_xref="FLYBASE:FBgn0000253"
protein_id="AAF58543.1"
/db_xref="GI:7303487"
/translation="MADOLTEEOIAEFKFAFLFDKDGDTITTKELGTVMRSLGONPTEAFLODMINEVDADNGTIDPEFLTMARKKDDTSEEIRARFVPDKDNGFISAAELRHVMTNGEKLTDPEVDMEIRADIDGQVNEEFVTMTSK"
join(21254..21464,21584..>21600)
/gene="CG13166"
/product="CT32407"
/db_xref="FLYBASE:FBan0013166"
/db_xref="FLYBASE:FBgn0040758"
/db_xref="FLYBASE:FBgn0040758"
/evidence=not_experimental
21254..>21600
/gene="CG13166"
/map="48F5-48F6"
/db_xref="FLYBASE:FBan0013166"
/db_xref="FLYBASE:FBgn0040758"
/evidence=not_experimental
join(21254..21464,21584..21600)
/gene="CG13166"
/note="CG13166 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0013166"
/db_xref="FLYBASE:FBgn0040758"
/evidence=not_experimental
protein_id="AAF58540.1"
/db_xref="GI:7303484"
/translation="MSERSGGAVTGGHRAVSPTPGSGSKATATPCRSRPPDIALAVPPPRNCILAI PMSLSLOFANGPGTKELHNP"
join(34383..34566,36219..36635,37304..37431,37529..37772,37837..37951,38022..38163,38232..38305,38481..>38679)
/gene="CG13165"
/product="CT32406"
/db_xref="FLYBASE:FBan0013165"
/db_xref="FLYBASE:FBgn0033707"
/evidence=not_experimental
34383..>38679
/gene="CG13165"
/map="48F5-48F8"
/db_xref="FLYBASE:FBan0013165"
/db_xref="FLYBASE:FBgn0033707"
/evidence=not_experimental
join(34383..34566,36219..36635,37304..37431,37529..37772,37837..37951,38022..38163,38232..38305,38481..38679)
/gene="CG13165"
/note="CG13165 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0013165"
/db_xref="FLYBASE:FBgn0033707"
/evidence=not_experimental
/evidence=not_experimental

```

/db_xref="id="AA58539.1"
/db_xref="GI:7303483"
/translation="MFRKNFREGASGGGGKLOSKANRTKRSDMGLVOOPPELHV
KTHLFPSPRPGVDGEERCSALSGIPSTPLSTAPSLAPGNLPYELDMPLVDR
PSVSLMRSSSGTGECSNSNSNSNSLNLLOOTOGOMOOLOOOLOOOQMO
OOOOLHATATATGADATVATATTLTLLAAATFOVNFEGGANSCTRAPEIS
LATLDDECTITVHSVDRPLPAGIPIPAISAGKATSPLEVSQPKCLIVHLSIP
LFAIDPVMSPTLAHLLKORSTIAELVTEVOQANVAGMILLANMKSSEEPFISVYL
INTLOTDPSTFYAGLRVSSLSKFEPEKALVTAHTLDLYSEVAATCRPPLVPSNEVG
SRKSGDKMYLLVCCEALDKDISLAFLIPALRRALCGYGTLYRPIQAF"
complement(join(39898, 40041, 40104, 40842, 40899, 41276,
41341, 41560, 41620, 41902, 41964, 42200, 43133, 43167))
/gene="CG8850"
/product="CT25422"
/db_xref="FLYBASE:FBan0008650"
/db_xref="FLYBASE:FBgn0033708"
complement(<39898, 43167)
/gene="CG8850"
/map="48F5-48F8"
/db_xref="FLYBASE:FBan0008650"

```

gene

Query Match

Best Local Similarity 64.0%; Score 16; DB 3; Length 266133;
Matches 16; Conservative 100.0%; Pred. No. 40;

Mismatches 0; Indels 0; Gaps 0;

Qy 10 gcatcgcacgcctcgt 25
|||||

Db 158787 GCATTTCGATCCTCGT 158772

RESULT 35

NTPECL/c 1389 bp DNA PLN 29-JUL-1996
LOCUS N.tabacum gene for pectate lyase.
DEFINITION X61102
ACCESSION X61102.1 GI:15981
VERSION X61102.1 GI:15981
KEYWORDS pectate lyase.
SOURCE common tobacco.
ORGANISM Nicotiana tabacum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
1 (bases 1 to 1389)

REFERENCE
AUTHORS Lonsdale, D.M.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-1991) D.M. Lonsdale, Cambridge Laboratory, Centre
for Plant Science Research, Colney Lane, Norwich, Norfolk NR4 7UJ,
UK

2 (bases 1 to 1389)
REFERENCE
AUTHORS Rogers, H.J., Harvey, A. and Lonsdale, D.M.
TITLE Isolation and characterization of a tobacco gene with homology to
pectate lyase which is specifically expressed during
microsporogenesis
JOURNAL Plant Mol. Biol. 20 (3), 493-502 (1992)
MEDLINE 93043039
COMMENT See X61101 for N.tabacum pectate lyase mRNA (partial).
FEATURES
Location/Qualifiers

1. 1389
/organism="Nicotiana tabacum"
/strain="Samsun"
/db_xref="taxon:4097"
/dev_stage="mature pollen"
/clone_1lb="lambda-EMBL3"
/clone="G10"
/number=1
/number=1
prim_transcript <1..>1389
CDS join(1..577,676..976,1074..1389)
/codon_start=1
/product="pectate lyase"
/protein_id="CAA43414.1"
/db_xref="GI:19982"

```

/db_xref="SMISS-PROT:P40972"
/translation="MDYRIRISVFLLVLTTPAALTTPATNIPROLSNKKYKPCRA
ENADIKWCRCOPMAENRORWADDALEGSNAIGKIGRIYVYTDNSDDVVPKPGT
LRYVLCCKEPLMTIEGKNMKIKLSRELIVSNKTIIGRGFNHIONGAGIKIGASANI
LISNLRHNTVPTPGILLRESEEDVGLRGSDEEDGSIIPESHIDWIIDHLSMRAPTGL
IDAAASTNITISNCHETDEHEKVLFGANDHYVLDMDKTTTLYVNHPRRLDORMRRC
RFGFHLVNDYTHMERVAIGSSGATIIISQGNRFIAEDDELLVKEVTYREKLTASVAE
KGRRC"
join(<1..577,676..976,1074..>1389)
578..675
/number=1
676..976
/number=2
977..1073
/number=2
1074..>1389
/number=3
exon
BASE COUNT 427 a 305 g 387 t
ORIGIN

```

Query Match

Best Local Similarity 60.0%; Score 15; DB 8; Length 1389;
Matches 15; Conservative 100.0%; Pred. No. 1,6e+02;

Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatcgcacgcctcgt 15
|||||

Db 162 GCATCGCCAGCATTT 148

RESULT 36

NTPECL/c 1713 bp DNA PLN 29-JUL-1993
LOCUS N.tabacum gene for pectate lyase.
DEFINITION X67158 S47561
ACCESSION X67158.1 GI:19907
VERSION X67158.1 GI:19907
KEYWORDS pectate lyase.
SOURCE common tobacco.
ORGANISM Nicotiana tabacum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
1 (bases 1 to 1713)

REFERENCE
AUTHORS Lonsdale, D.M.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-1992) D.M. Lonsdale, Cambridge Laboratory, Centre
for Plant Science Research, Colney Lane, Norwich Norfolk NR4 7UJ,
UK

2 (bases 1 to 1713)
REFERENCE
AUTHORS Rogers, H.J., Harvey, A. and Lonsdale, D.M.
TITLE Isolation and characterization of a tobacco gene with homology to
pectate lyase which is specifically expressed during
microsporogenesis
JOURNAL Plant Mol. Biol. 20 (3), 493-502 (1992)
MEDLINE 93043039
COMMENT
FEATURES
Location/Qualifiers

1. 1713
/organism="Nicotiana tabacum"
/strain="Samsun"
/db_xref="taxon:4097"
/germline
/dev_stage="mature pollen"
/clone_1lb="lambda EMBL3"
/clone="G10"
/number=1
/number=1
prim_transcript 108..1713
exon 108..901
mRNA join(108..901,1000..1300,1398..1713)
transl_peptide 325..399
CDS join(325..901,1000..1300,1398..1713)
/codon_start=1
/product="pectate lyase"

Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgcccagcattc 16
|||||
Db 1743 CATGCCAGCATTTT 1729

RESULT 39

PSEIAAHM/c

LOCUS

DEFINITION P_syringae tryptophan 2-monooxygenase (iaam) and indoleacetamide
hydrolyase (iaah) genes, complete cds.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

The iaah and iaam genes are part of an operon that is borne on a plasmid, pIAA, in oleander strains of the pathogen. There appears to be no genetic transformation of host tissue by P.savastanoi. No primow boxes were found, though potential ribosome binding sites are located at positions 435-439 and 2164-2168. Expression of iaah depends upon the promoter for iaam.
Draft entry and computer-readable sequence for [1] kindly submitted by T.Yamada, 14-JAN-1986.
Draft entry and computer-readable sequence for [1] kindly submitted by T.D.Gaffney, 28-JUN-1990.
In press] kindly submitted by J. Bacteriol.
Location/Qualifiers
1..3677
/organism="Pseudomonas syringae"
/sub_species="savastanoi"
/db_xref="taxon:317"
19..47
/note="indoleacetic acid operon promoter"
54..>3537
/note="indoleacetic acid mRNA"
82..312
/note="promoter proximal ORF"
/transl_table=1
/codon_start=1
/protein_id="AA25851.1"
/db_xref="GI:151290"
/translation="MPGRTPAANMLDKAMVAVARIROKPOAAVPAALLQAPHMETVW
STTKTECHCPAPYTPYTOASASASAVSAY"
454..2127
/note="tryptophan 2-monooxygenase (EC 1.13.12.3)"
/codon_start=1
/transl_table=1
/protein_id="AA25852.1"
/db_xref="GI:151291"
/translation="MYDHNPSIDILYDGPFLKCEMTGGISYASGTPPRVAIV
GAGISGLVATELLRAGVDVLYESRDRGGVWQVDVOTRPRYIAAGMARRPS
ATGLFYLKFGISTSTPPDDGVDTLEHYRKRHYHMPACKKPEPLFRVYEGMSI.
LSEGYLLEGGSLVAPLDTAMKLSGRLENAIAWQMLVFPDCSYNAIVLFTGRH
PFGGDWAPPEPELFGSLGSGGLFVFOAGFETELRMVINGYOSDRLIPDGISS
LAARLDQSTFDGKALRDVCSFVCRISREAEKIIQTPEGEORVDFRVTIVSSNRAM

FEATURES

Source

misc_feature

mRNA

CDS

1..5618
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="h11838"
/clone_lib="p1uescriptII SK plus"
/tissue_type="brain"
1..2935
/gene="K1A1057"
<1..2935
/gene="K1A1057"
/codon_start=2

OMIHLDSESESLRDVAVARETHLTGSSKFLTRTKFMINKKLPPTOSGLVIG
VYCLDYOPEDEEGHGVLLSYTWEDDOKMIAAMPDKRTKROVYLVDDAAIHPFASIL
LPVGDYERYVLIHDMLEDDPHSACAFPLNPGEDVYSORLPFOPMTANSNKTGTLYL
AGCSGFPAGNIEGAVOTALNSACAVLRSTGOLSKGNPLDCINASYRY"
2170..3537
/note="indoleacetamide hydrolyase"
/codon_start=1
/transl_table=1
/protein_id="AA25853.1"
/db_xref="GI:151292"
/translation="MHTITLESICQALADGETIAAELERRALDTEARLARLNCPTRE
GDVASFGEADHAKGTPLMGMEVSEFQNDICVGLPLTATGRMSGFVSPQDAIVSO
LRAIGAVAGAKNNMHLSPGVTSLNPMGTGNVPAVPGACGSSGSAAVAGVLP
LSVGTIDGSGIRIPAAFCGIGTGFPTTGRSTGIIIPVSHTKDCVGLTRGDAEFL
YGLSGKQSEFPLSRTPACRIGLPVMSWNSLDEVEERACVNAISLRKTFETEDD
ADIVELNQTLPFPLYEPFADLAQSLISGKMGHGHHPAIVDDANVAGIINHILGE
GAIRKPVYLSLQNGELKRRKMBELRARNIELLGYPTVPCRPVPHLDHADREFFSQA
IRNTDLSNMLPSITIPVGESERLPVGSFALRGDRDALLSRVSAIDVIGVFRKY
LPHTT"

CDS

BASE COUNT 813 a 917 c 1046 g 901 t
ORIGIN 1 bp upstream of EcoRI site.

Query Match 60.0%; Score 15; DB 1; Length 3677;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgcccagcattc 15
|||||
Db 1729 GCATGCCAGCATTTT 1715

RESULT 40
AB028980/c
LOCUS
DEFINITION Homo sapiens mRNA for KIAA1057 protein, partial cds.
ACCESSION AB028980
VERSION
KEYWORDS
SOURCE
Homo sapiens brain cDNA to mRNA, clone_11b:p1uescriptII SK plus
clone:h11838.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (sites)
Kikuno,R., Nagase,T., Ishikawa,K., Hirose,M., Miyajima,N.,
Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
Prediction of the coding sequences of unidentified human genes.
XIV. The complete sequences of 100 new cDNA clones from brain which
code for large proteins in vitro
DNA Res. 6 (3), 197-205 (1999)
99397452
2 (bases 1 to 5618)
Ohara,O., Nagase,T. and Kikuno,R.
Direct Submission
Submitted (17-JUN-1999) to the DDBJ/EMBL/GenBank databases. Osamu
Yana, Kazusa DNA Research Institute, Laboratory of DNA Technology,
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:cdna@kazusa.or.jp, Tel:+81-438-52-3913,
Fax:+81-438-52-3914)
Location/Qualifiers
1..5618
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="h11838"
/clone_lib="p1uescriptII SK plus"
/tissue_type="brain"
1..2935
/gene="K1A1057"
<1..2935
/gene="K1A1057"
/codon_start=2

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES
Source
1..5618
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="h11838"
/clone_lib="p1uescriptII SK plus"
/tissue_type="brain"
1..2935
/gene="K1A1057"
<1..2935
/gene="K1A1057"
/codon_start=2

LEIATIPYTHSGVLSALAMDKAQAKVAAGIPIVAGERNVNRKFDFTSEHPDQPVV
 VKRVRGSSFCVYIVKEDQSHPPQIITTSSEWFGQVMERYIHERELTGLDEAL
 GTEVPLDGHNEFYDAKYAGGSKHVPAGISPTIKYQITQLAVMAHQAIGCRVSK
 SDFRYDDRESEDEGVIMLEVNTQPGMTPTSLVPEMAAHAGRSFGDLVSMWVEDASCIK
 "

gene
 6894 .8198
 /gene="AGR_C_3790"
 6894 .8198
 /gene="AGR_C_3790"
 /note="HYPOTHELICAL METABOLITE TRANSPORT PROTEIN IN
 THEF-KDCK INTERGENIC REGION"
 /codon_start=1
 /transl_table=11
 /product="AGR_C_3790p"
 /protein_id="AAK87840.1"
 /db_xref="GI:15157222"
 /translation="MTDAVSRYSTAGNSNAVKTNSPARVLTAIVGTTFEPDEYV
 ATAALVFPALFPPNSDPTTLAFAFESIAFARPLGAVVFGYGRVGRKATVA
 ALLMGVSTVVVIGLPTVETAGVLAFLALCRRGGGGLGEMGAGVLLATENAPR
 KRWYGMFPGALPGVGLFLSSGVEMILLHPMSOELLISWGRIPPVASITITAVGMV
 RLSTETPDFOKAIPEERVAVPTALFERSKRSIVGTFVALATFVLEITIGVYL
 YNWKYKIPRLDALEVOVLGSLVIGLIPVAGKLAERFGRREVLTITLIGLFSPL
 PSLMTGGGSLFVFAALAMLMGMTYGLIGTALAAPPTRVRYTCSSITFMAGIFGA
 SLAPYATWLVQVNGMGVGYVLCISALITLACIFLSRKEV"
 complement(8268..9119)
 /gene="AGR_C_3792"
 complement(8268..9119)
 /gene="AGR_C_3792"
 /note="hypothetical protein PA5487 (imported) -
 pseudomonas aeruginosa (strain PAO1)"
 /codon_start=1
 /transl_table=11
 /product="AGR_C_3792p"
 /protein_id="AAK87841.1"
 /db_xref="GI:15157223"
 /translation="MOQPLARLDGTLISMGLMKRPTEMSSVSATQARRRLGPLE
 FSVRIACLVITIVTCTGGTLLADGTLVDFNGIMLSVTLGGAAVFTVLICVL
 NAREVQILVDSHERLHLSHTDALTLGSLNRLGIYAACAIGSDYCAVAFDIDFKAN
 DGYSHLDVLIASVARRIKHPDPAPHAVALRGEEFVVOETSPALFQMGSRVRA
 IETPVAVODLRITVITISIGVAFRGDAEIPDKVHNNDLALYRAKGGRRVRCMSGV
 ERROVAA"

BASE COUNT 1876 a 2805 c 2653 g 1894 t

ORIGIN

Query Match 60.0%; Score 15; DB 1; Length 9228;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 agcatttcgaccc 23
 |||||||||||||

Db 8578 AGCAATTCGATCTC 8564

RESULT 42
 AE004320/c DNA
 LOCUS
 DEFINITION Vibrio cholerae chromosome I, section 228 of 251 of the complete
 chromosome.
 ACCSSION AE004320 AE003852
 VERSION AE004320.1 GI:9657082
 KEYWORDS
 SOURCE
 ORGANISM
 Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 1 (bases 1 to 10205)
 Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwin, M.L.,
 Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,
 Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
 Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragol, I.,
 Sellers, P., McDonald, L., Uitterback, J., Fleischmann, R.D.,
 Nierman, W.C. and White, O.
 DNA sequence of both chromosomes of the cholera pathogen Vibrio

JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS

Cholerae
 Nature 406 (6795), 477-483 (2000)
 2 (bases 1 to 10205)
 Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwin, M.L.,
 Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,
 Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
 Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragol, I.,
 Sellers, P., McDonald, L., Uitterback, J., Fleischmann, R.D.,
 Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,
 Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
 Direct Submission
 Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 Location/Qualifiers
 1. 10205
 /organism="Vibrio cholerae"
 /strain="N16961"
 /serotype="O1"
 /db_xref="taxon:666"
 /chromosome="I"
 /note="biotype: El Tor"
 128..1639
 /gene="VC2501"
 128..1639
 /gene="VC2501"
 /note="similar to SP:P1648 GB:X15130 PID:1054725
 PID:43309 PID:537102; Identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="aminopeptidase A/1"
 /protein_id="AAF95643.1"
 /db_xref="GI:9657083"
 /translation="MEFSYSGSPKORSACIVGVFPPRRLSPVAPDLKISQCYIS
 SLRRLDLEGRGKGMILLHOVGVYSERLLVLCCKEELEROKETIOTITNLE
 TSGMEVVCITPELHYKGRDTYKAVQAVATKDGLYTIDQKSVKPEIRLRLKLVN
 VPTRRRLNIGERATITIGLATISGVACKDGLGMPNINAPVLAQARRLDVDEIST
 TTITGEEEMERKGMASVYLAVCGRSNEISMVETEKGNPDEAPRTVVGGLFEDSG
 GISTLRPGDEEMKYMCAASVFCGTAKIAIKLGLPLVIGLAGCEMPCSNLRVPG
 DILTTMSGQVTEVLTDAEGRVLCQDVLTYVERFPEPCVVAVATLTGACVATLGHHS
 AVMSHNPLAHELVNVAEISSDRARPLPLAEVYHQLCPSPADMANITGRCGATTA
 CFLSKFAKKYMAHLDIAGTAKKSAAGSTGRPVSLVQFLNRSGLDAAE"
 1723..2172
 /gene="VC2502"
 1723..2172
 /gene="VC2502"
 /note="similar to SP:P28905 SP:P11649 GB:L04574 GB:X15130
 GB:Z14155; Identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="DNA polymerase III, chi subunit"
 /protein_id="AAF95644.1"
 /db_xref="GI:9657084"
 /translation="MPTATFYILSRDSQAATPEGLAQVLYLTFRHFAAGAGARIYLOO
 DKPAEQLAEERFMOIEPQFLAHNVGSGPSSGATIGVGVSSPSNRRLAINTLADN
 OTTFARFTEVDEVPQAEENKQIARERYRLYRAGYQLOTVETIQY"
 2266..5127
 /gene="VC2503"
 2266..5127
 /gene="VC2503"
 /note="similar to SP:P07118 GB:J03497 GB:X05891 PID:147906
 PID:43302; Identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="valyl-tRNA synthetase"
 /protein_id="AAF95645.1"
 /db_xref="GI:9657085"
 /translation="MEKTVNPSTIEODLYKTWEQGFPRPHGTSKDAYSIMIPPNV
 TGLSHMGHAFODITMDTLRCORMKKNTLMOYGTGHAGIAPOMVYERKIAAEGETK
 HDYGRDAFIDIKIMWKAESGTTKOLRIGASVDMDRKREFTTDDGFTAVQDEVYRL
 YKDDLIRGKRLVNMWPKLHTAISDEVEKKEFGIMHVRPLAGGVATYADGKDYIV
 VATTREPMTLGDITGVAVNPEPRYKDLIGKEILLPIYGRRIPIVSGEADMEKGTGCV

KITPAHNDYEVGKRLHQLPMLNIFPDANI RDAEVEFNSNGEASNAVYETIPAKYOG
 MERPARAKIAVEFEELGLQETIKRDLTVYVGGDRGVVLEPMITQWYVRAVILAKP
 AVEAVENGDIOPFVGNENYFMSMRIOQWCSROLIMGHRIIPAWDEGVNFGVN
 BEEYRANNTIAADVALRODDVDLDTWSSALWTEKGTPELKVFPHTVLTG
 FDIIEVAVARIKMTHTFCDEDEGKAOPVRYVGLJEDENGDKMSKSGVLDPI
 DMTDIDLESIVAKRTGNMPOLAATKENTRTENGIEATGTSRLTAKMAAST
 GRIDIMDKMRLEGKRNFCNKLMAASRYLVNTEEDGCGFAAGALEYSLADKMAESF
 ELAKENENGHIDNFRIMANTLEYEIMNOCBMYLETPVLMKGTEAOQARTRL
 ITVEKTLRLAHPIYITTEIMOSVPLVGDGTIMQALPOYDVANONEDRL
 TEWAKTPTSTRNRAVEDINPKPLFVLMKAAEODAAKLEANKPYLVSLAKESIR
 VLADEATPACATALVCKSELMITMACLIQKDAELDLAKIEIAKTOGEIARIEGKIGN
 EGVAKAPEAVITREKRLAGVQEAIVKLEQKATIAL"
 /gene="VC2504"
 complement(5205, .6182)
 /note="similar to GB:ia2023 SP:p45250 PID:1007767
 PID:1221703 PID:1205791: identified by sequence
 similarity: putative"
 /transl_table=11
 /codon_start=1
 /product="2-hydroxyacid dehydrogenase family protein"
 /protein_id="AAF95646.1"
 /db_xref="GI:9657086"
 /translation="MSLPTSLPTVVPIDRATTRHISLPALPREHHLEADCEPOQ
 VVERLLADIVYTNKVLTRKMLIQPLRLKLIATISATGTNVDLPACDLNITACNVQ
 GYATRSVEHVYAMFALRNLIYGHNDIAGEGORHQCFFTHPIGIDLAGSTMGIT
 GSGAGQATNANLALGMHVLIAERKQCEVCRDGTSEFQVLAQSDVLSLHCPLTDET
 RNIITSEALQAKNPALLINRGSLGVDEQALVLDLKRQLAGAGVDFVSEAPADMN
 PLIANRDLNLIITPHVAVGSDSSIQQLATLLINDISAFMGEAKNRV"
 /gene="VC2505"
 complement(6222, .6959)
 /note="similar to SP:p39219 GB:U00096 PID:1786244:
 identified by sequence similarity: putative"
 /codon_start=1
 /transl_table=11
 /product="ribosomal large subunit pseudouridine synthase
 A"
 /protein_id="AAF95647.1"
 /db_xref="GI:9657087"
 /translation="MAMRENPKPDPMLDIVIEDEHIIANKPSGLISVGRVHYD
 SMKARLVDCPPVKYVRLDMSTGSLIFAKGHSALAKOFOVRLTKIYAVWG
 VMEONGEVNLPICDMPKPKOKVCFEGGRSLTRYOVKREKFTYVRLPIPGHS
 HOLRVHOSLGHPIVDDLYAPPEAVEAFRLQILAAELCFYHPSHMLRTLFPVCFD
 YPEAEOLIQHDPPEKLPDYKLSOD"
 /gene="VC2506"
 complement(7056, .9965)
 /note="similar to GB:U00096 PID:1786245: identified by
 sequence similarity: putative"
 /codon_start=1
 /transl_table=11
 /product="RNA polymerase-associated protein Hepa"
 /protein_id="AAP95648.1"
 /db_xref="GI:9657088"
 /translation="MSFALGQWNIPTDESGLGIVVALDARTVLTAMFASEENRYA
 RSDAPYRVTFVNGVDVDSQWGLQVEVDEGVYTYLGTVDVEEGVALRETEL
 SNOIREKPODKLFAQOIDRMNVLRYALTYOQOHSFPMGLGSMAGLIPHLY
 IAHVEGRHRAPIRLDEVGLTIEGQITGQVLETRERELIIVVPELLOHOMVE
 MMRENNHSLPDEBERVEAFSEADNPFEQOYVLCSDLRKSRQRESHFARLILAD
 LVDEAHHLDMHEKSPREVOYIEALAEQPPVLLITATPEOIGRSHFARLILAD
 RPYDEAFVAREEYAFVADAVTALFSGEKLSDPAKNAIGFVRNHLLELPSYTTSM
 ASDEELIAROEELIDNLMDRHGTVLENTRAIAGFVRNHLLELPSYTTSM
 RVAGMLGKLTPEARAKMLYPEIIOFEGESSWMPDSRVNMLLEKAKRSKI
 LVASRASTALLOEALREREGIRATVHEGMSIIEERKAAVFAQEGEAGVYLSE
 IGSEGRNFAFNCPTGRAYVADAEERLPIIAGGEGEELVIEESAKNTKLKLSQ
 RMPOEGIMAFNCPPTGRAYVADAEERLPIIAGGEGEELVIEESAKNTKLKLSQ
 EVGRDLRLHNSNGEAKAOIADYIATDDDTJNYPALSIPLAIGIHODRGENALV
 VTPAEHMAVPSYGLPYEGATITFDRTALSRDMFTISNEHNVQCGDILLSEVCG
 TCAVSLKNKALPVGTLLLELVVYVDAQAKRSISKRLPVSPIRLIMAKRNDLSQ

Query Match	Score 15: DB 1: Length 10205;
Best Local Similarity 100.0%; Pred No. 1,6e+02;	
Matches 15: Conservative 0; Mismatches 0; Indels 0; Gaps 0;	

Qy	8	cagcattcgatcct 22
Db	6278	CAGCATTTCGATCCT 6264

RESULT	43
AE003860	
LOCUS	
DEFINITION	Xylella fastidiosa 9a5c, section 6 of 229 of the complete genome.
ACCESSION	AE003860 AE003849
VERSION	AE003860.1 GI:9104830
KEYWORDS	
SOURCE	
ORGANISM	Xylella fastidiosa 9a5c. Bacteria: Proteobacteria; gamma subdivision; Xanthomonas group; Xylella.
REFERENCE	
AUTHORS	1 (bases 1 to 10593) Simpson, A.J., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M., Alvarenga, R., Alves, L.M., Araya, J.E., Bala, G.S., Baptista, C.S., Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Briones, M.R., Bueno, M.R.P., Camargo, A.A., Camargo, L.E.A., Carraro, D.M., Carrier, H., Costa-Neto, C.M., Coutinho, L.L., Cristofani, M., Dias-Neto, C., El-Dorri, H., Faciocani, A.P., Ferreira, A.J., Ferreira, V.C., Franco, M., Furlan, L.R., Fraga, J.S., Frasca, S.C., Franco, M.C., Goldman, M.H., Gomes, S.L., Gruber, A., Ho, P.L., Hohnselt, J.D., Junqueira, M.L., Kemper, E.L., Kitaajima, J.P., and Marino, C.L. The genome sequence of the plant pathogen Xylella fastidiosa. The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing and Analysis Nature 406 (6792), 151-157 (2000)
JOURNAL	
MEDLINE	
PUBMED	
REFERENCE	
AUTHORS	2 (bases 1 to 10593) Simpson, A.J.G., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M., Alvarenga, R., Alves, L.M.C., Araya, J.E., Bala, G.S., Baptista, C.S., Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Briones, M.R.S., Bueno, M.R.P., Camargo, A.A., Camargo, L.E.A., Carraro, D.M., Carrier, H., Costa-Neto, C.M., Coutinho, L.L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorri, H., Faciocani, A.P., Ferreira, A.J.S., Ferreira, V.C.A., Ferro, J.A., Fraga, J.S., Frasca, S.C., Franco, M.C., Frohme, M., Furlan, L.R., Garnier, M., Goldman, G.H., Goldman, M.H.S., Gomes, S.L., Gruber, A., Ho, P.L., Hohnselt, J.D., Junqueira, M.L., Kemper, E.L., Kitaajima, J.P., Krieger, J.E., Kuramae, E.E., Laigret, F., Lambais, M.R., Leite, L.C.C., Lemos, E.G.M., Lemos, M.V.B., Lopes, S.A., Lopes, C.R., Machado, J.A., Machado, M.A., Madeira, A.M.B.N., Madira, H.M.F., Marino, C.L., Marques, M.V., Martins, E.A.L., Martins, E.M.F., Matsukuma, A.Y., Menck, C.F.M., Miracca, E.C., Miyaki, C.Y., Monteiro-Vitorello, C.B., Moon, D.H., Nagai, M.A., Nascimento, A.L.T.O., Netto, L.E.S., Niani Jr., A., Nobrega, F.G., Nunes, L.R., Oliveira, M.A., de Oliveira, M.C., de Oliveira, R.C., Palmeri, D.A., Paris, A., Peixoto, B.R., Pereira, G.A.G., Pereira, J.R., H.A., Pesquero, J.B., Quaggio, R.B., Roberto, P.G., Rodrigues, V., de M. Rosa, A.J., de Rosa Jr., V.E., de Sa, R.G., Santelli, R.V., Sawasaki, H.E., da Silva, A.C.R., da Silva, F.R., da Silva, A.M., Silva Jr., M.A., da Silveira, J.F., Silvestri, M.L.Z., Siqueira, W.J., de Souza, A.A., de Souza, A.P., Tenenzi, M.F., Truffi, D., Tsai, S.M., Tsunako, M.H., Vallada, H., Van Sluys, M.A., Verjovski-Almeida, S., Vettore, A.L., Zaio, M.A., Zatz, M., Zedler, J. and Setubal, J.C. Direct Submission Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and

Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
13083-970, Brazil

FEATURES

source

Location/Qualifiers
1..10593

/organism="Xylella fastidiosa 9a5c"
/db_xref="taxon:160492"

/clone="9a5c"

/complement(277..1356)

/gene="XF0061"

/complement(277..1356)

/note="similar to SP|P07674 (percent identity: 28 %/query alignment coverage: 64.9 %/subject alignment coverage: 65.1 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="transcriptional repressor"

/protein_id="AAF82874.1"

/db_xref="GI:9104831"

/translation="MLLOTLELDRAATAMISFGNRIRKPTQEOQLAVAFSGHIE
IDPELIDLEKORSKDNPGFTTESIMEIGNDRKRGQHEPVLKRNKKPGRYLMAV
ERRMLACKAVDIKKAVIRELDDOVRVQAEANTQRENLQLEIAVALRDKERLGT
LEKVAENKKGITNWSERKINELEVETNSLSQAVSEGITADITAINDLVLEKVNCT
AAKTLIDAKVDPGIMRKRYREKLQNAKPTTNEKNATDSRMQESKIDPEGETHIQ
NISNOKPATYASAPLPDAQVALAKOMETLCTEVSALSKILHQKQDVISALGV
LQNVQAQALFRTTHVLASLANSK"

/complement(1331..1657)

/gene="XF0062"

/complement(1331..1657)

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF82875.1"

/db_xref="GI:9104832"

/translation="MYRSIFORSNSTINETVKHCKWSDLNHDIDTTPUPSTNLI
LEMDFNKTNEICAEINTSPLOSIRPRPEQHNIFILPHLMVQHLKMQARIA
VVTDIV"

/complement(2427..3152)

/gene="XF0063"

/complement(2427..3152)

/note="similar to GI|5921548 (percent identity: 37 %/query alignment coverage: 93.4 %/subject alignment coverage: 95.7 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="complete protein F"

/protein_id="AAF82876.1"

/db_xref="GI:9104833"

/translation="MKDAVNSNHLIDPTWCDELRLMLDSYCLACNEPQHOGTLCAT
COLLIPRENAELRCAPLKEGILHPTGCPCLDPPPLETAAHAFLYMPLINKLSR
FKTHODLACGHLADNMMAAQAHPPOALVPSLHHGRLERTGYDQALELRHNSR
LQPCPLALRRVATLPOSOLNANKRRSNLHNAFIDILPHLPAVALVDVMTGTTH
AAARALRHGTVIDAMICARVP"

/complement(6494..7495)

/gene="XF0064"

/complement(6494..7495)

/note="similar to SP|P26601 (percent identity: 52 %/query alignment coverage: 86.8 %/subject alignment coverage: 99.7 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="biotin synthase"

/protein_id="AAF82877.1"

/db_xref="GI:9104834"

/translation="MSIIRKDWIAEELHALFDLSLPCLLYVRAASVHRQHFPAEIGV

gene

CDS

4281..4382

/gene="XF0065"

/complement(4281..4382)

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF82878.1"

/db_xref="GI:9104835"

/translation="MATLVDVSPMYDGVNLDSDSGERRKVEFS"

/gene="XF0066"

/complement(4486..5673)

/note="similar to SP|O34731 (percent identity: 33 %/query alignment coverage: 63.3 %/subject alignment coverage: 96.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAF82879.1"

/db_xref="GI:9104836"

/translation="MSLINHKNGVSCNCSTCSRDILNLRVRVSGCNRPYHLSY
LALSMCFTRFCSFLATIVALLTACSSAPRSASRLSSVSLDISLPIRLGALGG
GAAGFAHIGVLMLEANGLASVAGTSAGSVCALYASGMDSKTQTKAYMDESN
IDLRPEGGLVOGOKLONVYNEIYKGRTEKLNKPFGLATRLDQRTMVRNGV
EAVRASCSTPGVEPYTIGAHYHVDGGITSPVPVADVQGLADFDVADISSKAKN
PGLVGTVMOSTSIMGHLEAEKLRANVTHPKVLDIGSTDFDGRNRAIIEGERATI
AVIPETREKLAQLOOORIOARKTANOSIEAVESYRRCNSRTREKLRKCKCTK
SIL"

/complement(6138..6377)

/gene="XF0067"

/complement(6138..6377)

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF82880.1"

/db_xref="GI:9104837"

/translation="MSCKEIFEYIVIKQCMLEPFPAQMGVMGAIWCADRYPIVLIS
RCVLSTYVWGCVRFESVILMPMCWLGVMIVAF"

/complement(6494..7495)

/gene="XF0068"

/complement(6494..7495)

/note="similar to SP|P26601 (percent identity: 52 %/query alignment coverage: 86.8 %/subject alignment coverage: 99.7 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="hydroxybenzoate octaprenyltransferase"

/protein_id="AAF82881.1"

/db_xref="GI:9104838"

/translation="MLINQHEHNYNHNHLSNFORPLNPSNSPVSMAVYFRTSAI
TILHMNRNLDPIYKLABDGPVGFLLMLPMTWALMAGVPPMWTLCVPTGTIML
TRSGACVINDYTRWDPDHVERIRIRLVGAVSPNALMFAITMLIAFGLITFQNL
LTVLISVGLFLAMTYPYLKRYYTLPOVYGLAFEGGIPAFAGIAGKVPQOAWLLYV
ANILMTAYDTWCAMVDRODDIKGAKSTALIFADGLTYOCVLYTLMLFTCLVLR
ATLSHTYWSILAAVVALIGYFIARRRRTACFRFPMNNVGMITIFAGIALATTH"

Query Match

60.08; Score 15; DB 1; Length 10593;

gene
 CDS
 /codon_start=1
 /transl_table=11
 /product="AGR_C_1547p"
 /protein_id="AAK86654.1"
 /db_xref="GI:15155834"
 /translation="MPILYTKASATGAGNAGKSESDVLTDTLVPEKLGSGDAK;IT
 NEPQLEAGYSACFGLKAVAGKHVKVLPEDTLYATATGIGPREDGFGFGLVTLKV
 NIPGLERKAEELVAAAHVCPYSHAMRSTSEVPSVA"
 7902..8363
 /gene="AGR_C_1550"
 7902..8363
 /gene="AGR_C_1550"
 /note="hypothetical protein"
 /codon_start=1
 /transl_table=11
 /product="AGR_C_1550p"
 /protein_id="AAK86655.1"
 /db_xref="GI:15155835"

gene
 CDS
 /translation="WGPGLSRPSCLOKEICYLISKERICAMNIIHAKTAMDPAQGRV
 TKAVIAAERGLNARAPADILGVAPVSRMRRLDFLEGAKSFEELVILIVFRS
 LDAITGDEAVAKNMLRNHNTALDAVPAERLTITIGLIDVLSYDARRAPV"
 8323..9084
 /gene="AGR_C_1551"
 8323..9084
 /gene="AGR_C_1551"
 /note="hypothetical protein PA1030 (imported)"
 /codon_start=1
 /transl_table=11
 /product="AGR_C_1551p"
 /protein_id="AAK86656.1"
 /db_xref="GI:15155836"

gene
 CDS
 /translation="WSSPITWTFDYLSEKRAASKFWRLVEAQHOVSTMKLVPTVERO
 SLDDITLSTSKRPPPCAGDYLLATPFPGAYPHGSRFRAGYEGEYVAAKVE
 TALAEMAFYRLIFAESGTPLPANPADYSAPARLATDAALDITDPLNRDETAMTD
 LTYNEPCQALAEQARLAKVEALIVRSYRDAACINIALISPKFAEKAPVERMSWRIR
 LSKTVOALCEFFPRKTRGPAVSPAGPRLASLIG"
 complement(9108..10622)
 /gene="AGR_C_1553"
 complement(9108..10622)
 /gene="AGR_C_1553"
 /note="1553"
 /codon_start=1
 /transl_table=11
 /product="AGR_C_1553p"
 /protein_id="AAK86657.1"
 /db_xref="GI:15155837"

BASE COUNT 2028 a 3320 c 3297 g 2221 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 1; Length 10866;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 cgcacacatttcgat 19
 |||||||
 Db 3005 CGCCACGATTTCGAT 3019

RESULT 45
 AE004795 10978 bp DNA BCT 30-AUG-2000
 LOCUS
 DEFINITION Pseudomonas aeruginosa PA01, section 356 of 529 of the complete

ACCESSION AE004795 AE004091
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Pseudomonas aeruginosa.
 Pseudomonas aeruginosa
 Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 Pseudomonas.
 1 (bases 1 to 10978)
 Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warren, P.,
 Hickey, M. J., Brinkman, F. S., Hufnagle, W. O., Kowalik, D. J., Lagrou, M.,
 Garber, R. L., Goltzy, L., Tolentino, E., Westbrock-Wadman, S., Yuan, X.,
 Brody, L. K., Coulter, S. N., Folger, K. R., Kas, A., Lapidis, K. A., Smith, K. A.,
 Smith, K. K., Spencer, D., Wong, G. K., Wu, Z., and Paulsen, I. T.
 Complete genome sequence of Pseudomonas aeruginosa PA01, an
 opportunistic pathogen
 Nature 406 (6799), 959-964 (2000)
 20437337
 2 (bases 1 to 10978)
 Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warren, P.,
 Hickey, M. J., Brinkman, F. S., Hufnagle, W. O., Kowalik, D. J.,
 Lagrou, M., Garber, R. L., Goltzy, L., Tolentino, E.,
 Westbrock-Wadman, S., Yuan, X., Brody, L. K., Coulter, S. N.,
 Folger, K. R., Kas, A., Lapidis, K. A., Lim, R. M., Smith, K. A., Spencer, D. H.,
 Wong, G. K., S., Wu, Z., Paulsen, I. T., Reizer, J., Salier, M. H.,
 Hancock, R. E. W., Lory, S., and Olson, M. V.
 Direct Submision
 Submitted (16-MAY-2000) Department of Medicine and Genetics,
 University of Washington Genome Center, University Of Washington,
 Box 352145, Seattle, WA 98195, USA

FEATURES

source
 1..10978
 /organism="Pseudomonas aeruginosa"
 /strain="PA01"
 /db_xref="taxon:287"
 /db_xref="GI:9949931"
 complement(132..449)
 /gene="PA3762"
 complement(132..449)
 /gene="PA3762"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK07149.1"
 /db_xref="GI:9949932"
 /translation="MKLCFVYPESHLDVYKQAVFAGGCGICAGVDSQWGLGCGF
 RPLDSQPYLGQVQVEHVAEMKVELVADLIIHASKALKAAHPYETPAYEWRLLD
 MVF"
 complement(914..4810)
 /gene="putL"
 /note="PA3763"
 complement(914..4810)
 /gene="putL"
 /codon_start=1
 /transl_table=11
 /product="phosphoribosylformylglycinamide synthase"
 /protein_id="AAK07150.1"
 /db_xref="GI:9949933"

/translation="MLTIRGAPALSAFRHGLLEOLTOHVPAYTGHAFAFRHADV
 ALTDEQVILARLILKYGSPVOEPSGRPLVVPFGTISPMSSKASDIARNGAKTI
 DLREGIAYVVOGLSESDAQOAAARLHDMTOVLVDRLEGAELESHQPRITLAVD
 VLGGRALAEKANVELGLAIEDEIDYILKSPGLRNPDELIMFAQNSHCHK
 IFNARSVIDGOADOKSLFGMKIKNYEMNREGVLSAYKDNAAVYGVHAGVFFPDQTR
 EYASRPPDILMKVETHNHPAIAFPAGSTSGGEIRDEGATGGAAPKALNGET
 VSNOLPGFROPHEVPRGKERLVTAUDLIDNBGLGAAAFNNEGRPALGYFRPEO
 KIATPHREEVYGHKPLMLAGMGNTIRDEHVOKGETISVGAKLIVLGPPMLIGLGGA
 ASSMTCASSADIDFASVDRNDPEMRCEVIDRCWOLGERNPISFIDHVGAGISN
 ALPELINDGGRGRFELRAVPNDPEMSPLEIWCNSOREYVLSVAAPFEPKATCE
 RERCPEAVVGEATEIROLTVADSHFENKPVMDPLLEVLIGKAPMHRARREAGDGF
 DAAGLEIEMSERVLRHPAVASKSPILTIGDRTITGLVARDOVNGVPOVPVADCAVTA
 TSFDVYTGEMAMGERTPLALDAPASGMAIGETVNTAAARVCKLSDIKLSANMA
 AAGHCEDEALITYTAVAGMEICPELGITITPVGKDSMSKTRWODGEDKSVTSVSL
 IYGVFPAVDAROSTITPOLRLKGETDLILIDGRCNKLGSIIAOVHGRKGRVAVP
 VDAEDLKAFFAVIOGLINADGHILAYHDSRDLITSLYEMAFAGHCGVELMLDLAD

```

SREELAAVLFSEELGAVIQVREGATPEVLAQFSAGLIDDCVAVIGOPVNGCYELINLYN
GEIYVSAORRILLOINSETSYOIORLNRNADCAOEFDALLDEDNPLSIRKISYDVND
DIAPYIKKGVKPVVALLREGVNGVEMAAEDRAGFAAIDVMSDILLAGRDLPALF
KGIYAGCGFSYGVILGEGEMAKSILFNARARQFOAFPAKDSFALGVNGCGOMNS
LHEILPTETEPHPHVNRNSQDFEAFVARNMOVOSSESTFLOGMAGSRPLPILAIAGEHA
EFSEELALLEADISGCVALRFDVMDHGVTTATYIPANPQSGRIGITGLSSROCRVLTIMR
HPERVFRVONSMRBDMDOEDGWLMEFRNARVWD"
/gene="PA3764"
5211..6569
/gene="PA3764"
5211..6569
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG07151.1"
/db_xref="GI:9949934"
/transl_table="MOKRGVLRVITRNSPATYPODRNGETGEFEYELAKREAEERLVEL
KITRDNLDLILYAOISREGCPALAAAGLTPGRREDAVRSYSHYLDVTPOLIYRNGOO
RPTREDLVGRIMVILKSSHAEDLAEUKOYPELKYEBSDAVEVVDLLRMVQDID
LTIYDSNELANNQYFPVNVVAFDFEGEGRALMPLGGDDDSLMEVNAFLDQAKKGS
LLORLKDRYGVHVDLVGVAVTFAOHLQOQLPRYESHFQKSGQLDPTDMLAAISY
OESLMOPGATSKTGVRGLMLLTNRQAQMGVSNRLDPKOSIQGSKYFVOIRSELPS
IKEDPSWPFALAAVNIAGLHEDAKMAEKELNPKMKLDVYKMLPRIADKOMYAKIR
YGIARGETVHFVONVRKYVDILTMVTOPOMESSQIAESGLHLPGVVKTRFEEDSGDE
KL"
complement(6583..7167)
/gene="PA3765"
complement(6583..7167)
/gene="PA3765"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG07152.1"
/db_xref="GI:9949935"
/transl_table="MANDPIQSFYCFDGTGTRRLITPAALPEVALAKQAVAGAGAPLL
VFDNATGRSVDDIRGSSDPLARLVAGEROPAAEAPRCGRKLCVAREVTLIP
RHWMIAOPGASVALRIKRVLEARRSQGRARQOEARVHYHMNMAAGDLDPFEATR
ALFPADREGFAERLIGWPTDIVERHAAMLASDGS"
complement(7293..8549)
/gene="PA3766"
complement(7293..8549)
/gene="PA3766"
complement(7293..8549)
/codon_start=1
/transl_table=11
/product="probable aromatic amino acid transporter"
/protein_id="AAG07153.1"
/db_xref="GI:9949936"
/transl_table="MSDTRVOQYODHAGAEAVDSOLEYKRLTFLEAVAMIVGTNIGA
GYLSMAYASRKAGFMPLLMLAVAGLFTTISMLYSETALRTTHNOLSLAORYGS
FGMAFELISVAVNSGALIAVMSGCKILSAFEGISPALGSVLFPIAGVLYLGISA
IGKEGETISIMKSMILITVAATLLANDTEPARLLDGDWYVWPVNTIAVPCESAQYI
VPEMARGFSHAPERLKPAVITGMITFVLITPLSVLIALTGLENOSEVATLAKSKAL
GEWAFPTANTFALCAMLTSYWGIGSELTNITFDKFRKLGEPNPKTRILVLAIVALP
FVLAISGLVFNALYFAGFSGVILSIMPIILMRSARRNGDEPAWKCGWIAHPAIQ
AVIVITYLASAAYAIASAMNLLPAGW"
complement(8938..9486)
/gene="PA3767"
complement(8938..9486)
/gene="PA3767"
complement(8938..9486)
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG07154.1"
/db_xref="GI:9949937"
/transl_table="MRYTKLISNGKCYPTKYKVRPIIDRSRDPHPMREALAEKRAA
LGEVPGAVLVREGELIGRFNRPISSHDSHAEMLAIRMAAAGNTRLPSTILY
TLPPCSMCSGLVHARIORLVGTVPEKSGAVESRGRFEOHLNHRVVEGCVLAE
CSOALSAFFRAREAAKRPRED"
complement(9477..10868)
/gene="PA3768"
complement(9477..10868)
/gene="PA3768"
complement(9477..10868)
/codon_start=1
/transl_table=11
/product="probable metallo-oxidoreductase"
/protein_id="AAG07155.1"
/db_xref="GI:9949938"
/transl_table="MTFTROYLGGLAVVGLAGAGARLMLARPOVAQETVIELIA
APLDLEIVGFSSPALAYGQCQGVELRAKQGEMLRVRTNLTDEPTTIHMGIRLPI
EMGVYIISOPVQGESEFIVQETQODAGSYWYHPLMSSDGLGLVPLIIEBERP
TGRHHRKVICLKTWHVDEGQAFTPSPVPROAAREGRGYSTINGKHVPTIDLPAGOI
VRYRLNVNTVITYRLNPLNGEARKIYATIDGHPVEPREGGVYIGRMLTELKVPK
AGTELSDRGVPVRLATIRSVASAEAPAGDMPPLPANPYSEPDLMNAEKIGRFEWVG
AMSDTSGKNPSPFWQINGKAMEGGEERHNAAPLAKLEGOSTIFELNMAQYOHPI
HLHGMAFKVLDSBRDIIIPFTDYLLGKNETARVALVADNGLMFMCHVTDHMETG
LMGTIAGEAWCG"
BASE COUNT 1876 a 3658 c 3644 g 1800 t
ORIGIN
Query Match 60.0%; Score 15; DB 1; Length 10978;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 atgcgacacattcg 17
Db 9239 ATCGCACACATTTCG 9253

```

Search completed: December 26, 2001, 12:52:48
Job time: 8890 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:16:47 ; Search time 368.09 Seconds
(without alignments)
58.228 Million cell updates/sec

Title: us-09-396-196f-7

Perfect score: 25
Sequence: 1 gcatcgccagcattcgcctcgt 25

Scoring table: OLIGO-NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 9

Total number of hits satisfying chosen parameters: 32850

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : N_Geneseq_1101:*

1: /SID22/gcgcdata/geneseq/geneseqn/NA1980.DAT.*
2: /SID22/gcgcdata/geneseq/geneseqn/NA1981.DAT.*
3: /SID22/gcgcdata/geneseq/geneseqn/NA1982.DAT.*
4: /SID22/gcgcdata/geneseq/geneseqn/NA1983.DAT.*
5: /SID22/gcgcdata/geneseq/geneseqn/NA1984.DAT.*
6: /SID22/gcgcdata/geneseq/geneseqn/NA1985.DAT.*
7: /SID22/gcgcdata/geneseq/geneseqn/NA1986.DAT.*
8: /SID22/gcgcdata/geneseq/geneseqn/NA1987.DAT.*
9: /SID22/gcgcdata/geneseq/geneseqn/NA1988.DAT.*
10: /SID22/gcgcdata/geneseq/geneseqn/NA1989.DAT.*
11: /SID22/gcgcdata/geneseq/geneseqn/NA1990.DAT.*
12: /SID22/gcgcdata/geneseq/geneseqn/NA1991.DAT.*
13: /SID22/gcgcdata/geneseq/geneseqn/NA1992.DAT.*
14: /SID22/gcgcdata/geneseq/geneseqn/NA1993.DAT.*
15: /SID22/gcgcdata/geneseq/geneseqn/NA1994.DAT.*
16: /SID22/gcgcdata/geneseq/geneseqn/NA1995.DAT.*
17: /SID22/gcgcdata/geneseq/geneseqn/NA1996.DAT.*
18: /SID22/gcgcdata/geneseq/geneseqn/NA1997.DAT.*
19: /SID22/gcgcdata/geneseq/geneseqn/NA1998.DAT.*
20: /SID22/gcgcdata/geneseq/geneseqn/NA1999.DAT.*
21: /SID22/gcgcdata/geneseq/geneseqn/NA2000.DAT.*
22: /SID22/gcgcdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	839	22	AA162941 Human genomic DNA
2	25	100.0	1041	20	AA01303 E. coli biotin syn
3	25	100.0	1084	10	AA091329 E. coli Bio B gene
4	25	100.0	1121	7	AA060496 Sequence encoding
5	25	100.0	5872	15	AA062386 Biotin-biosynthesis
6	15	60.0	5662	22	AA161089 Human polynucleoti
7	15	60.0	5773	21	AA076014 Human ORFX ORF1569
8	15	60.0	5876	22	AA159303 Human polynucleoti
9	14	56.0	473	22	AA03765 Drosophila sulfata
10	14	56.0	487	21	AA074884 Human ORFX ORF439
11	14	56.0	549	21	AA093834 Cat flea hindgut a

C	12	14	56.0	614	22	AA054422 Mammalian vestibul
C	13	14	56.0	943	22	AA148556 Human fascin DNA f
C	14	14	56.0	1240	22	AA148555 Human fascin DNA f
C	15	14	56.0	1344	21	AA256381 Escherichia coli f
C	16	14	56.0	1449	22	AA148554 Human fascin DNA f
C	17	14	56.0	1619	22	AA148553 Human fascin DNA f
C	18	14	56.0	1934	22	AA148552 Human fascin DNA f
C	19	14	56.0	3069	22	AA148551 Human fascin DNA f
C	20	14	56.0	4100	22	AA148550 Human fascin DNA f
C	21	14	56.0	4199	22	AA148549 Murine multidrug r
C	22	14	56.0	4313	14	AA038950 Mouse multidrug re
C	23	14	56.0	5319	19	AAV35235 R. prowazekii S-1a
C	24	14	56.0	13055	22	AA148621 Human fascin DNA f
C	25	14	56.0	16951	22	AA148620 Human fascin DNA f
C	26	14	56.0	16951	22	AA148622 Human fascin DNA f
C	27	14	56.0	143068	21	AA021105 Human low adenosin
C	28	14	56.0	143068	21	AA021105 Human low adenosin
C	29	14	56.0	143068	21	AA021105 Human low adenosin
C	30	14	56.0	143068	21	AA021105 Human low adenosin
C	31	14	56.0	149412	21	AA035151 Human adenosine re
C	32	14	56.0	152740	21	AA035151 Human adenosine re
C	33	13	52.0	51	21	AA076687 Human adenosine re
C	34	13	52.0	51	21	AA076687 Human adenosine re
C	35	13	52.0	61	20	AA019606 Human gene express
C	36	13	52.0	300	20	AA019606 Human gene express
C	37	13	52.0	322	21	AA019606 Human gene express
C	38	13	52.0	435	22	AA114896 Plant microstereili
C	39	13	52.0	435	22	AA114896 Plant microstereili
C	40	13	52.0	435	22	AA114896 Plant microstereili
C	41	13	52.0	447	17	AA027962 Probe #4669 used t
C	42	13	52.0	447	17	AA027962 Probe #4669 used t
C	43	13	52.0	591	21	AA054077 Hepatitis C virus
C	44	13	52.0	597	22	AA054077 Hepatitis C virus
C	45	13	52.0	606	21	AA054077 Hepatitis C virus
C	46	13	52.0	629	21	AA054077 Hepatitis C virus
C	47	13	52.0	709	22	AA069558 Arabidopsis thalia
C	48	13	52.0	714	20	AA069558 Arabidopsis thalia
C	49	13	52.0	725	22	AA069558 Arabidopsis thalia
C	50	13	52.0	750	20	AA069558 Arabidopsis thalia
C	51	13	52.0	765	22	AA069558 Arabidopsis thalia
C	52	13	52.0	974	21	AA035170 Human gene express
C	53	13	52.0	974	21	AA035170 Human gene express
C	54	13	52.0	985	21	AA046920 Human gene express
C	55	13	52.0	996	21	AA046920 Human gene express
C	56	13	52.0	1007	19	AA0296443 Human gene express
C	57	13	52.0	1038	21	AA012052 Arabidopsis thalia
C	58	13	52.0	1065	20	AA012052 Arabidopsis thalia
C	59	13	52.0	1110	18	AA084197 Arabidopsis thalia
C	60	13	52.0	1114	21	AA084197 Arabidopsis thalia
C	61	13	52.0	1179	21	AA015961 Human protein clon
C	62	13	52.0	1226	21	AA036239 Arabidopsis thalia
C	63	13	52.0	1263	9	AA082205 Nad B gene encodin
C	64	13	52.0	1287	21	AA03673 Human peroxisome a
C	65	13	52.0	1291	21	AA03673 Human peroxisome a
C	66	13	52.0	1302	22	AA066075 C glutamicum codin
C	67	13	52.0	1313	21	AA066075 C glutamicum codin
C	68	13	52.0	1340	21	AA066075 C glutamicum codin
C	69	13	52.0	1357	21	AA066075 C glutamicum codin
C	70	13	52.0	1364	20	AA066075 C glutamicum codin
C	71	13	52.0	1404	20	AA066075 C glutamicum codin
C	72	13	52.0	1419	20	AA066075 C glutamicum codin
C	73	13	52.0	1419	20	AA066075 C glutamicum codin
C	74	13	52.0	1419	20	AA066075 C glutamicum codin
C	75	13	52.0	1425	22	AA066075 C glutamicum codin
C	76	13	52.0	1430	22	AA066075 C glutamicum codin
C	77	13	52.0	1443	22	AA066075 C glutamicum codin
C	78	13	52.0	1467	22	AA066075 C glutamicum codin
C	79	13	52.0	1475	22	AA066075 C glutamicum codin
C	80	13	52.0	1481	22	AA066075 C glutamicum codin
C	81	13	52.0	1497	20	AA066075 C glutamicum codin
C	82	13	52.0	1515	20	AA066075 C glutamicum codin
C	83	13	52.0	1515	20	AA066075 C glutamicum codin
C	84	13	52.0	1566	22	AA070984 C. glutamicum SRT

```

C 85 13 52.0 1606 21 AAC41788 Arabidopsis thaliana
C 86 13 52.0 1641 22 AAH61477 C. glutamicum codin
C 87 13 52.0 1650 17 AAT28576 Bacterial antibody
C 88 13 52.0 1654 21 AAT15174 Trichoderma reesei
C 89 13 52.0 1694 21 AAC37317 Arabidopsis thaliana
C 90 13 52.0 1724 11 AA003361 The nada nucleotid
C 91 13 52.0 1724 11 AA003363 The nada nucleotid
C 92 13 52.0 1727 22 AAC87201 Alcohoh and/or ald
C 93 13 52.0 1737 19 AA29053 C. glutamicum SRP
C 94 13 52.0 1764 22 AAT71128 P. putida K12440-a
C 95 13 52.0 1791 22 AAT61039 Sul gene containing
C 96 13 52.0 1804 11 AA004673 Pseudomonas sp hea
C 97 13 52.0 1815 22 AAT6413 Human secreted pro
C 98 13 52.0 1887 22 AAH19221 DNA encoding a Bac
C 99 13 52.0 1950 20 AAH86149 E. coli growth and
C 100 13 52.0 1992 22 AAH84533

```

ALIGNMENTS

RESULT 1

AAI62941 standard; DNA: 839 BP.

22-OCT-2001 (first entry)

Human genomic DNA SEQ ID NO 269.

Human; noctropic; neuroprotective; cytosolic; dermatological; virucide;
 Immunosuppressant; anti-inflammatory; anti-HIV; antibacterial; vulvar;
 antiparasitic; hepatotropic; cerebroprotective; antihistaminic;
 antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
 antiparasitic; cardiant; immune disorder; cardiovascular disorder;
 neurological disease; infection; nephrotropic; gene therapy; vaccine;
 ds.

Homo sapiens.

WO20015449-A1.

02-AUG-2001.

17-JAN-2001; 2001WO-US01346.

31-JAN-2000; 2000US-0179065.
 04-FEB-2000; 2000US-0180628.
 19-MAY-2000; 2000US-0205515.
 07-JUL-2000; 2000US-0216880.
 14-JUL-2000; 2000US-0218290.
 14-AUG-2000; 2000US-0225447.
 01-SEP-2000; 2000US-0229343.
 06-SEP-2000; 2000US-0230437.
 08-SEP-2000; 2000US-0231243.
 25-SEP-2000; 2000US-0234997.
 29-SEP-2000; 2000US-0236387.
 13-OCT-2000; 2000US-0239337.
 08-NOV-2000; 2000US-0246476.
 08-NOV-2000; 2000US-0246477.
 08-NOV-2000; 2000US-0246525.
 08-NOV-2000; 2000US-0246526.
 08-NOV-2000; 2000US-0246528.
 17-NOV-2000; 2000US-0249210.
 17-NOV-2000; 2000US-0249211.
 17-NOV-2000; 2000US-0249214.
 01-DEC-2000; 2000US-0250160.
 01-DEC-2000; 2000US-0250391.
 05-DEC-2000; 2000US-0251030.
 05-DEC-2000; 2000US-0251968.

05-DEC-2000; 2000US-0256719.
 06-DEC-2000; 2000US-0251479.
 08-DEC-2000; 2000US-0251989.
 11-DEC-2000; 2000US-0251990.
 2000US-0254097.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI: 2001-476225/51.

Novel plasma membrane associated proteins useful for diagnosing,
 treating, preventing and/or prognosing disorders related to the
 proteins, including cancer, immune response and neuronal disorders

Example 2; SEQ ID NO 269; 532pp + Sequence Listing; English.

The invention relates to novel genes (AAI62752-AAI62961) and proteins
 (AAM42347-AAM42415) useful for preventing, treating or ameliorating
 medical conditions e.g. by protein or gene therapy. The genes are
 isolated from a range of human tissues disclosed in the specification.
 The nucleic acids, proteins, antibodies and (ant)agonists are useful
 in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 and ovarian cancer and other cancers of the adrenal gland, bone, bone
 marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 colitis; (c) cardiovascular disorders such as myocardial ischemia;
 (d) wound healing; (e) neurological diseases such as cerebral anoxia and
 epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 and parasitic infections.

Note: The sequence data for this patent did not form part of the
 printed specification, but was obtained in electronic format directly
 from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other;

Query Match 100.0%; Score 25; DB 22; Length 839;
 Best Local Similarity 100.0%; Pred. No. 3e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 gcatcgccagcattgcattcgtcgt 25
 Db 122 gcatcgccagcattgcattcgtcgt 146

RESULT 2

AAAX01303 standard; DNA: 1041 BP.

AAAX01303;

12-APR-1999 (first entry)

E. coli biotin synthetase (BioB) coding sequence.

DAP aminotransferase; diaminopelargonic acid; transgenic plant;
 biotin synthetase; biotin production; vitamin H; BioB; ss.

Escherichia coli.

US5869719-A.

09-FEB-1999.

30-APR-1997; 97US-0846338.

30-APR-1997; 97US-0846338.
 08-MAR-1995; 95US-0401068.

PA (NOVS) NOVARTIS FINANCE CORP.
 XX Patton DA:
 XX
 XX
 DR WPI: 1999-152902/13.
 DR P-PSDB: AAW73906.
 XX
 XX Transgenic plants with high biotin levels - transformed with DNA
 PT encoding di:amino-pelargonic acid amino-transferase or biotin
 PT synthase
 XX
 PS Example 2: Column 37-40: 34pp; English.
 XX
 CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.
 XX
 SQ Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other.

Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 3e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcatgcgcagcattcgcattcgcctcgt 25
 ||||||||||||||||||||||||||||
 Db 90 gcacgcgcagcattcgcattcgcctcgt 114

RESULT 3
 ID AAN91329 standard; DNA: 1084 BP.
 XX
 AC AAN91329;
 XX
 DT 15-FEB-1990 (first entry)
 XX
 DE E.coli Bio B gene.
 XX
 KM E.coli: Bio B gene: biotin.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT CDS 24..1064
 FT /*tag=a
 FT
 XX
 PN GB2216530-A.
 XX
 PD 11-OCT-1989.
 XX
 PE 17-MAR-1989; 89GB-0006210.
 XX
 PR 22-MAR-1988; 88GB-0006804.
 PR 17-MAR-1989; 89GB-0006210.
 XX
 PA (UKAG-) UK MIN. AGRIC. FISH.
 XX
 PI Pearson BM, McKee RA;
 XX
 DR WPI: 1989-295085/41. P-PSDB P91392
 XX
 PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.
 XX
 PS Table 3; page 33-4; 52pp; English.
 XX
 CC The gene can be used in a plasmid for expression of enzymes of the biotin

CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
 CC Lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.
 XX
 SQ Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 3e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcatgcgcagcattcgcattcgcctcgt 25
 ||||||||||||||||||||||||||||
 Db 113 gcacgcgcagcattcgcattcgcctcgt 137

RESULT 4
 ID AAN60496 standard; DNA: 1121 BP.
 XX
 AC AAN60496;
 XX
 DT 17-OCT-1991 (first entry)
 XX
 DE Sequence encoding biotin synthesising enzyme.
 XX
 KM Biotin synthetic enzyme: E.coli; desbiohiotin: ds.
 XX
 FH Key Location/Qualifiers
 FT CDS 42..1082
 FT /*tag= a
 FT
 XX
 PN Jp61149091-A.
 XX
 PD 07-JUL-1986.
 XX
 PE 24-DEC-1984; 84JP-0272605.
 XX
 PR 24-DEC-1984; 84JP-0272605.
 XX
 PA (NIPS) NIPPON SODA KK.
 XX
 DR WPI: 1986-216622/33.
 DR P-PSDB: AAP60536.
 XX
 PT Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
 XX
 PS Disclosure: Page 534; 23pp; Japanese.
 XX
 CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing desbiohiotin.
 XX
 SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 3e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcatgcgcagcattcgcattcgcctcgt 25
 ||||||||||||||||||||||||||||
 Db 131 gcacgcgcagcattcgcattcgcctcgt 155

RESULT 5
 ID AAO62386 standard; DNA: 5872 BP.
 XX

```

AC  AAO62386;
XX
XX  16-NOV-1994 (first entry)
XX
DE  Biotin-biosynthesis genes contg. plasmid PB030A-15/9.
XX
XX  Biotin: expression: enterobacteria; vitamin H; synthesis;
XX  plasmid: PB030A-15/9; biob; bioc; bioc; bioc; bioc;
XX  promoter plac: biotin synthase; KAPA synthase;
XX  8-amino-7-oxononanoate synthase; pimeoyl-CoA; DTB synthase;
XX  dehydrobiotin synthase; DAPA synthase;
XX  S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
XX  seborrhoea; dermatitis; ds.
XX
OS  Escherichia coli DSM498.
XX
XX  Key
XX  Location/Qualifiers
XX  1..96
XX  /tag= a
XX  /function= "promoter plac"
XX  /evidence= EXPERIMENTAL
XX  23..28
XX  /tag= b
XX  /standard_name= "promoter plac"
XX  45..50
XX  /tag= c
XX  /evidence= EXPERIMENTAL
XX  /standard_name= "promoter plac"
XX  105..109
XX  /tag= d
XX  /evidence= EXPERIMENTAL
XX  /standard_name= "biob RBS no. 9"
XX  117..1157
XX  /tag= e
XX  /product= "biotin synthase"
XX  /evidence= EXPERIMENTAL
XX  /gene= "biob"
XX  /number= 1
XX  1141..1146
XX  /tag= f
XX  /standard_name= "bioc RBS"
XX  1154..2311
XX  /tag= g
XX  /EC_number= 2.3.1.47
XX  /product= "KAPA synthase"
XX  /evidence= EXPERIMENTAL
XX  /gene= "bioc"
XX  /number= 2
XX  /standard_name= "8-amino-7-oxononanoate synthase"
XX  2284..2288
XX  /tag= h
XX  /standard_name= "bioc RBS"
XX  2295..3050
XX  /tag= i
XX  /function= "involved in pimeoyl-CoA synthesis"
XX  /product= "protein"
XX  /gene= "bioc"
XX  /number= 3
XX  /tag= j
XX  /standard_name= "bioc RBS"
XX  3043..3753
XX  /tag= k
XX  /EC_number= 6.3.3.3
XX  /product= "DTB synthase"
XX  /evidence= EXPERIMENTAL
XX  /gene= "bioc"
XX  /number= 4
XX  /standard_name= "dehydrobiotin synthase"
XX  3712..3750
XX  /tag= l
XX  /note= "bioc substitution"
XX  3742..3746
XX  RBS

```

```

FT  /tag= m
FT  /standard_name= "bioc RBS"
FT  3750..5039
FT
FT  CDS
FT
FT  /tag= n
FT  /EC_number= 2.6.1.62
FT  /product= "DAPA synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "bioc"
FT  /number= 5
FT  /standard_name= "S-adenosyl-L-methionine: 8-amino-
FT  7-oxononanoate aminotransferase"
FT
FT  RBS
FT  5088..5093
FT  /tag= o
FT  /standard_name= "ORFI RBS"
FT  5098..5574
FT  CDS
FT
FT  /tag= p
FT  /function= "unknown, involved in biotin synthesis"
FT  /product= "protein"
FT  /evidence= EXPERIMENTAL
FT  /gene= "ORFI"
FT  /number= 6
FT  5583..5644
FT  /tag= q
FT  /standard_name= "rho-independent transcriptional
FT  terminator"
FT  5583..5605
FT  /tag= r
FT
FT  stem_loop
FT
FT  WO9408023-A.
FT
FT  14-APR-1994.
FT
FT  01-OCT-1993; 93WO-EP02688.
FT
FT  02-OCT-1992; 92CH-0003124.
FT  15-JUL-1993; 93CH-0002134.
FT
FT  (LONZ ) LONZA AG.
FT
FT  Birch O, Brass J, Fuhrmann M, Shaw N;
FT
FT  WPI; 1994-135587/16.
FT  P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
FT
FT  Biotechnological biotin prodn. using enterobacterial biotin-gene
FT  - providing vitamin H in high yield
FT
FT  Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
FT
FT  The sequence is derived from plasmid PB030A-15/9 contg. the
FT  biob, bioc, bioc and bioc genes responsible for biosynthesis
FT  of biotin, arranged in a transcription unit. Microorganisms
FT  contg. these DNA fragments or plasmids may be used in the prodn.
FT  of biotin. Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
FT  loss of appetite and tiredness.
FT
FT  Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;
FT
FT  Query Match
FT  Best local Similarity 100.0%; Score 25; DB 15; Length 5872;
FT  Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
FT
FT  QY 1 gcatgcgcagcattcgcctcctcgt 25
FT  DB 206 gcatgcgcagcattcgcctcctcgt 230
FT
FT  RESULT 6
FT  ID AAI61089/c
FT  XX AAI61089 standard; cDNA; 5662 BP.
FT  AC AAI61089;

```


Query Match 60.0%; Score 15; DB 21; Length 5773;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 catgccagcatc 16
 |||||||
 DB 4886 CATGCCAGCATTTTC 4872

RESULT 8
 AA159303/c
 ID AA159303 standard; cDNA; 5876 BP.
 XX AA159303;

DE 22-OCT-2001 (first entry)
 XX Human polynucleotide SEQ ID NO 1506.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukemia; ss.

OS Homo sapiens.

XX WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000MO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

(HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

DR WPI: 2001-442253/47.
 P-PSDB: AAM40147.

PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -

XX Claim 1: SEQ ID NO 1506; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukemias and
 CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX Sequence 5876 BP; 1704 A; 1165 C; 1288 G; 1719 T; 0 other;

Query Match 60.0%; Score 15; DB 22; Length 5876;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 catgccagcatc 16
 |||||||
 DB 5014 CATGCCAGCATTTTC 5000

RESULT 9
 AAD03765/c
 ID AAD03765 standard; cDNA; 473 BP.
 XX AAD03765;

DE 19-JUN-2001 (first entry)

DE Drosophila sulfatase (Dmsulf) EST AA391898 cDNA.

KW Sulfatase; Sulf-1; Sulf-2; cytostatic; vitruicide; antiinflammatory;
 KW degenerative disease; neural; renal; skeletal muscle; viral infection;
 KW metastasis; inflammation; cancer; EST; Expressed Sequence Tag; therapy;
 KW ss.

OS Drosophila sp.

XX WO200121640-A1.

PD 29-MAR-2001.

PF 22-SEP-2000; 2000MO-US26124.

PR 23-SEP-1999; 99US-0155738.

PA (UYPE-) UNIV PENNSYLVANIA.
 PA (ROYA-) ROYAL VETERINARY COLLEGE.

PI Emerson CP, Dhoot GK;

DR WPI: 2001-266062/27.

PT Novel Sulf-1 or Sulf-2 (members of subfamily of sulfatases) polypeptide
 PT useful for treating musculoskeletal, neural or renal degenerative
 PT disorder, and for inhibiting viral infection of cells -

XX Claim 2: Page 34; 59pp; English.

XX The present sequence is Drosophila sulfatase (Dmsulf) EST (Expressed
 CC Sequence Tag) AA391898 cDNA.
 CC The invention relates to Sulf-1 and Sulf-2 proteins and their
 CC corresponding cDNA molecules which are the members of subfamily of
 CC sulfatases. These sulfatase proteins are expressed in neural and muscle
 CC lineages in various species. Sulfatase proteins are useful for modifying
 CC growth properties of cells, preferably cancer cells, useful in the
 CC treatment of cancer and in the inhibition of metastases. Sulf-1 and
 CC Sulf-2 are useful in developing cells for transplant in the treatment of
 CC skeletal muscular degenerative diseases, neurodegenerative diseases, renal
 CC degenerative diseases and in initiation growth of healthy cells and to
 CC heal diseased cells in these disorders. Sulfatases are also useful for
 CC inhibiting infection of cells by viruses which utilize sulfated heparin
 CC proteoglycans for entry into cells, and for modulating recruitment of
 CC lymphocytes by cells to sites of inflammation. A functional embryonic
 CC technique is useful to functionally characterize members of Sulf-1 and
 CC Sulf-2 sulfatase gene subfamily, which is efficient and economical.
 XX Sequence 473 BP; 100 A; 135 C; 128 G; 108 T; 2 other;

Query Match 56.0%; Score 14; DB 22; Length 473;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 7 cagcattcgatcc 20
 |||||
 221 CCAGCATTTTCGATC 208

RESULT 10
 AAC74884
 ID AAC74884 standard; cDNA; 487 BP.
 AC AAC74884;
 DT 08-FEB-2001 (first entry)
 DE Human ORFX ORF439 polynucleotide sequence SFO ID NO:877.
 XX
 XX Human: open reading frame; ORFX: detection; cytoskeletal; hepatotropic;
 KW vulnery; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;
 KW antiviral; antibacterial; antifungal; antineumatic; antihypertensive;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX
 XX Homo sapiens.
 OS
 XX
 PN WO200058473-A2.
 PD 05-OCT-2000.
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 PI
 PI Shinkets RA, Leach M;
 DR WPI; 2000-602362/57.
 DR P-PDB; AAB40675.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 899-900; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in ABA40237 to ABA43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytoskeletal; hepatotropic; vulnery;
 CC antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antinflammatory; antibacterial; antiviral; antifungal; antineumatic;
 CC antihypertensive; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy

CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 487 BP; 122 A; 99 G; 129 G; 137 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 487;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 8 cagcattcgatcc 21
 |||||
 316 cagcattcgatcc 329

RESULT 11
 AAC93854/C
 ID AAC93854 standard; cDNA; 549 BP.
 AC AAC93854;
 DT 19-FEB-2001 (first entry)
 DE Cat flea hindgut and Malpighian tubule (HMT) cDNA, SFO ID NO:349.
 XX
 XX Cat flea: hindgut and Malpighian tubule nucleic acid; HMT;
 KW flea infestation; vaccine; antiparasitic; therapeutic target;
 KW diagnosis; detection; ss.
 KW
 KW Ctenocephalides felis.
 OS
 XX
 PN WO200061621-A2.
 PD 19-OCT-2000.
 PF 07-APR-2000; 2000WO-US09437.
 XX
 PR 09-APR-1999; 99US-0128704.
 XX
 PA (HESK-) HESKA CORP.
 PI
 PI Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;
 DR WPI; 2000-656323/63.
 XX
 PT Flea Malpighian tubule and head and nerve cord tissue derived nucleic
 PT acids useful for the prevention, diagnosis and treatment of flea
 PT infestations -
 XX
 PS Claim 26; Page 361; 964pp; English.
 XX
 CC The invention relates to novel cat flea (Ctenocephalides felis) nucleic
 CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
 CC or head and nerve cord (HNC) tissue. The invention also relates to the
 CC encoded proteins. The invention additionally encompasses expression
 CC constructs, recombinant viruses and recombinant cells comprising the
 CC nucleic acids of the invention, recombinant production of the proteins,
 CC antibodies against the proteins, a method of identifying inhibitors of
 CC the proteins, and compositions comprising the inhibitors for
 CC administration to an animal. The nucleic acids, and the proteins they
 CC encode may be used in the prevention, treatment and diagnosis of diseases
 CC associated with flea infestations. For example, the nucleic acids may be
 CC used to produce an HMT or HNC protein according to standard recombinant
 CC DNA methodology by inserting the nucleic acids into a host cell and
 CC culturing the cell to express the protein. The HMT and HNC nucleic acids
 CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect

CC and quantitate the presence of cat flea or other homologous nucleic acid
 CC sequences in samples. They may also be used to study the expression and
 CC function of the proteins and their role in metabolism. The HMT and HNC
 CC proteins may be used as antigens in the production of specific
 CC antibodies, and in assays to identify modulators (agonists and
 CC antagonists) of HMT and/or HNC protein expression and activity. The
 CC anti-HMT/HNC protein antibodies and antagonists may also be used to
 CC downregulate protein expression and activity. The antibodies may also be
 CC used as diagnostic agents for detecting the presence of flea polypeptides
 CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
 CC present sequence represents a cat flea HMT cDNA of the invention.
 XX
 SQ Sequence 549 BP; 155 A; 102 C; 130 G; 160 T; 2 other;

Query Match 56.0%; Score 14; DB 21; Length 549;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgcccagcattt 15
 |||||
 Db 421 CATGCCAGCATTT 408

RESULT 12
 AAS05422/c
 ID AAS05422 standard; DNA: 614 BP.
 XX

AC AAS05422;

DT 07-SEP-2001 (first entry)

DE Mammalian vestibular system geotactic behaviour modulator gene #22.

KM Mammalian vestibular system; invertebrate; geotactic behaviour; vertigo;
 KM graviperceptive disorder; motion sickness; labyrinthitis; syphilis; ds;
 KW Meniere's disease; acoustic neuroma; multiple sclerosis; epilepsy;
 KW trauma; infection of the middle ear; ototoxic agent exposure.
 XX

OS Drosophila melanogaster .

PN WO200140519-A2.

PD 07-JUN-2001.

PF 01-DEC-2000; 2000WO-US32639.

PR 02-DEC-1999; 99US-0168579.

PR 26-SEP-2000; 2000US-0669751.

PA (NEUR-) NEUROSCIENCES RES FOUND INC.

PI Greenspan RJ;

DR WPI; 2001-356159/37.

PT New isolated nucleic acid having mammalian vestibular system-modulating
 PT activity useful in the treatment of disorders such as motion sickness
 PT and vertigo -
 XX

PS Claim 59; Page 99; 179pp; English.

CC The sequences shown in AAS05401-AAS05661 represent DNA with mammalian
 CC vestibular system-modulating activity. The DNA sequences can be used in a
 CC method whereby a first and second strain of an invertebrate is obtained,
 CC and both are subjected to conditions in which the strains exhibit
 CC different geotactic behaviour. Genes that are differentially expressed in
 CC the first strain relative to the second strain are then identified.
 CC Mammalian genes having substantially the same nucleic acid sequence as
 CC these modulate the mammalian vestibular system. Compounds containing
 CC these genes are used to decrease the symptoms of graviperceptive
 CC disorders such as motion sickness, vertigo, labyrinthitis, Meniere's
 CC disease, acoustic neuroma, multiple sclerosis, syphilis, trauma,

CC infection of the middle ear, exposure to ototoxic agents and epilepsy.
 XX
 SQ Sequence 614 BP; 140 A; 176 C; 160 G; 138 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 614;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ccagcatttcgattc 20
 |||||
 Db 217 CCAGCATTTGATTC 204

RESULT 13
 AAH48556
 ID AAH48556 standard; DNA: 943 BP.
 XX

AC AAH48556;

DT 20-SEP-2001 (first entry)

DE Human fascin DNA fragment SEQ ID 6.

KM Fascin; regulatory sequence; human; dendritic cell; antiviral; tumor;
 KM antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
 KW immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
 KW Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
 KW autoimmune disease; transplant rejection; ds.
 XX

OS Homo sapiens.

PN WO200151631-A2.

PD 19-JUL-2001.

PF 12-JAN-2001; 2001WO-EP00362.

PR 13-JAN-2000; 2000DE-1001169.

PR 02-MAR-2000; 2000DE-1010188.

PA (RESK/) RESKE-KUNZ A.

PA (ROSS/) ROSS X.

PA (ROSS/) ROSS R.

PA (BROS/) BROS M.

PI Reske-Kunz A, Ross X, Ross R, Bros M;

DR WPI; 2001-451858/48.

PT New regulatory sequences from the fascin gene, useful for providing
 PT dendritic cell-specific expression of e.g. antigens, e.g. for
 PT vaccination against tumors and infections -
 XX

PS Claim 1c; Page 87; 117pp; German.

CC This invention describes novel regulatory sequences (A) derived from
 CC human fascin that provide specific expression in dendritic cells (DC) and
 CC which have antiviral, antibacterial, antifungal, antiparasitic,
 CC anti-allergic, neurological, immunomodulatory and apoptotic activity.
 CC (A) are used to regulate expression of antigens, immunoregulators,
 CC antinease sequences etc. in DC-specific fashion. Recombinant DNA, vectors
 CC and host cells that contain (A) are useful: (i) in vaccines against
 CC viruses, bacteria, fungi, parasites, tumors, allergens and plaques in
 CC Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of
 CC rejection. They can also be provide specific expression of antigens and
 CC immunoregulators in DC; for isolation and identification of cell factors
 CC and cis-elements from regulatory sequences that mediate DC-specific
 CC expression; to determine the degree of maturity of DC and to block
 CC transcription factors, by providing binding sites in DC. (A) provide
 CC DC-specific expression of nucleic acid under their control, allowing a
 CC more specific regulation of the immune response and eliminating the long

CC and laborious purification of DC (since a complete leucocyte population
CC may be transformed), including transformation in vitro. This sequence
CC represents a DNA fragment of human fascin which is described in the
CC invention.
XX
SO Sequence 943 BP; 135 A; 307 C; 340 G; 161 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 943;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcatcgccagcatt 14
111 gcatcgccagcatt 124
Db

RESULT 14
ID AAH48555 standard; DNA; 1240 BP.
XX
AC AAH48555;
XX
DT 20-SEP-2001 (first entry)
XX
DE Human fascin DNA fragment SEQ ID 5.
XX
KW Fascin; regulatory sequence; human; dendritic cell; antiviral; tumor;
KW antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
KW immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
KW Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
KW autoimmune disease; transplant rejection; ds.
XX
OS Homo sapiens.
XX
PN WO200151631-A2.
XX
PD 19-JUL-2001.
XX
PF 12-JAN-2001; 2001WO-EP00362.
XX
PR 13-JAN-2000; 2000DE-1001169.
PR 02-MAR-2000; 2000DE-1010188.
XX
PA (RESK/) RESKE-KUNZ A.
PA (ROSS/) ROSS X.
PA (ROSS/) ROSS R.
PA (BROS/) BROS M.
XX
PI Reske-Kunz A, Ross X, Ross R, Bros M;
XX
DR WPI; 2001-451858/48.
XX

PT New regulatory sequences from the fascin gene, useful for providing
PT dendritic cell-specific expression of e.g. antigens, e.g. for
PT vaccination against tumors and infections -
XX
XX Claim 1c: Page 87; 117pp; German.

CC This invention describes novel regulatory sequences (A) derived from
CC human fascin that provide specific expression in dendritic cells (DC) and
CC which have antiviral, antibacterial, antifungal, antiparasitic,
CC anti-allergic, neurological, immunomodulatory and apoptotic activity.
CC (A) are used to regulate expression of antigens, immunoregulators,
CC antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors
CC and host cells that contain (A) are useful: (i) in vaccines against
CC viruses, bacteria, fungi, parasites, tumors, allergens and plaques in
CC Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of
CC tumors, allergies, infections, autoimmune diseases and transplant
CC rejection. They can also be provided specific expression of antigens and
CC immunoregulators in DC; for isolation and identification of cell factors
CC and cis-elements from regulatory sequences that mediate DC-specific
CC expression; to determine the degree of maturity of DC and to block

CC transcription factors, by providing binding sites in DC. (A) provide
CC DC-specific expression of nucleic acid under their control, allowing a
CC more specific regulation of the immune response and eliminating the long
CC and laborious purification of DC (since a complete leucocyte population
CC may be transformed), including transformation in vitro. This sequence
CC represents a DNA fragment of human fascin which is described in the
CC invention.
XX
SO Sequence 1240 BP; 191 A; 364 C; 466 G; 219 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1240;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcatcgccagcatt 14
408 gcatcgccagcatt 421
Db

RESULT 15
ID AA256381/C
ID AA256381 standard; DNA; 1344 BP.
XX
AC AA256381;
XX
DT 17-MAR-2000 (first entry)
XX
DE Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:51.
XX
KW Flagellin; flag; antigen; detection; ds.
XX
OS Escherichia coli.
XX
PN WO9961458-A1.
XX
PD 02-DEC-1999.
XX
PF 21-MAY-1999; 99WO-AU00385.
XX
PR 21-MAY-1998; 98AU-0003634.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Reeves PR, Wang L;
XX
DR WPI; 2000-072598/06.
XX
PT Novel nucleic acid molecule useful for the detection of flagellated
PT bacterial strains in food, faeces, etc. -
XX
XX Claim 3; Page 225; 245pp; English.

CC AA256331 to AA256398 represent nucleic acid molecules (1) encoding all
CC or part of an Escherichia coli flagellin protein except a protein
CC expressed by E. coli H1, H7, H12 or H48 type strains. The present
CC invention also describes a method of detecting the presence of E. coli
CC of a particular H serotype in a sample, comprising specifically
CC hybridising a nucleic acid, preferably at least a pair, derived from a
CC flagellating gene, specific for a particular flagellin gene associated
CC with the H serotype, to any E. coli in the sample which contain the gene,
CC and detecting any hybridised molecules, identifying the presence of that
CC serotype in the sample. (1) are useful for: (1) detecting the presence
CC of E. coli of H serotype in a sample by hybridising at least one or a
CC pair of (1) to any E. coli in the sample and detecting the hybridised
CC nucleic acid molecules; and (2) for detecting the presence of both O
CC and H-serotypes of E. coli by hybridising at least one or a pair of (1)
CC to any E. coli present in the sample and detecting the hybridised
CC nucleic acid molecules. (1) is particularly useful for detecting the
CC combination of O and H antigen. Hybridised (1) when using at least one
CC (1) is detected by southern blot analysis and, when using a pair of (1),
CC is detected by polymerase chain reaction (PCR). AA256399 to AA256420
CC represent primers used in the exemplification of the present invention.

XX S0 Sequence 1344 BP; 362 A; 337 C; 322 G; 323 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1344;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatcgccagcatt 14
 |||||
 Db 590 GCATCGCCAGCATT 577

RESULT 16
 AAH48554
 ID AAH48554 standard; DNA; 1449 BP.

AC AAH48554;
 XX
 DT 20-SEP-2001 (first entry)

DE Human fascin DNA fragment SEQ ID 4.

XX Fascin: regulatory sequence; human; dendritic cell; antiviral; tumor;
 KM antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
 KM immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
 KM Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
 KM autoimmune disease; transplant rejection; ds.

OS Homo sapiens.

PN WO200151631-A2.

XX 19-JUL-2001.

PD 12-JAN-2001; 2001WO-EP00362.

PF 13-JAN-2000; 2000DE-1001169.

PR 02-MAR-2000; 2000DE-1010186.

XX (RESK/) RESKE-KUNZ A.

PA (ROSS/) ROSS X.

PA (ROSS/) ROSS R.

PA (BROS/) BROS M.

PI Reske-Kunz A, Ross X, Ross R, Bros M;

XX WPI: 2001-451858/48.

DR

XX

PT New regulatory sequences from the fascin gene, useful for providing

PT dendritic cell-specific expression of e.g. antigens, e.g. for

PT vaccination against tumors and infections

XX

XX

PS Claim 1c; Page 86-87; 117pp; German.

CC This invention describes novel regulatory sequences (A) derived from
 CC human fascin that provide specific expression in dendritic cells (DC) and
 CC which have antiviral, antibacterial, antifungal, antiparasitic,
 CC anti-allergic, neurological, immunomodulatory and apoptotic activity.
 CC (A) are used to regulate expression of antigens, immunoregulators,
 CC antisense sequences etc. In DC-specific fashion. Recombinant DNA, vectors
 CC and host cells that contain (A) are useful: (i) in vaccines against
 CC viruses, bacteria, fungi, parasites, tumors, allergens and plaques in
 CC Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of
 CC tumors, allergies, infections, autoimmune diseases and transplant
 CC rejection. They can also be provide specific expression of antigens and
 CC immunoregulators in DC; for isolation and identification of cell factors
 CC and cis-elements from regulatory sequences that mediate DC-specific
 CC expression; to determine the degree of maturity of DC and to block
 CC transcription factors, by providing binding sites in DC. (A) provide
 CC DC-specific expression of nucleic acid under their control, allowing a
 CC more specific regulation of the immune response and eliminating the long
 CC and laborious purification of DC (since a complete leucocyte population

CC may be transformed), including transformation in vitro. This sequence
 CC represents a DNA fragment of human fascin which is described in the
 CC invention.

XX S0 Sequence 1449 BP; 232 A; 429 C; 532 G; 256 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1449;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatcgccagcatt 14
 |||||
 Db 617 gcatcgccagcatt 630

RESULT 17
 AAH48553
 ID AAH48553 standard; DNA; 1619 BP.

AC AAH48553;

XX 20-SEP-2001 (first entry)

DE Human fascin DNA fragment SEQ ID 3.

XX Fascin: regulatory sequence; human; dendritic cell; antiviral; tumor;
 KM antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
 KM immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
 KM Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
 KM autoimmune disease; transplant rejection; ds.

OS Homo sapiens.

PN WO200151631-A2.

XX 19-JUL-2001.

PD 12-JAN-2001; 2001WO-EP00362.

PF 13-JAN-2000; 2000DE-1001169.

PR 02-MAR-2000; 2000DE-1010186.

XX (RESK/) RESKE-KUNZ A.

PA (ROSS/) ROSS X.

PA (ROSS/) ROSS R.

PA (BROS/) BROS M.

PI Reske-Kunz A, Ross X, Ross R, Bros M;

XX WPI: 2001-451858/48.

DR

XX

PT New regulatory sequences from the fascin gene, useful for providing

PT dendritic cell-specific expression of e.g. antigens, e.g. for

PT vaccination against tumors and infections

XX

XX

PS Claim 1c; Page 86; 117pp; German.

CC This invention describes novel regulatory sequences (A) derived from
 CC human fascin that provide specific expression in dendritic cells (DC) and
 CC which have antiviral, antibacterial, antifungal, antiparasitic,
 CC anti-allergic, neurological, immunomodulatory and apoptotic activity.
 CC (A) are used to regulate expression of antigens, immunoregulators,
 CC antisense sequences etc. In DC-specific fashion. Recombinant DNA, vectors
 CC and host cells that contain (A) are useful: (i) in vaccines against
 CC viruses, bacteria, fungi, parasites, tumors, allergens and plaques in
 CC Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of
 CC tumors, allergies, infections, autoimmune diseases and transplant
 CC rejection. They can also be provide specific expression of antigens and
 CC immunoregulators in DC; for isolation and identification of cell factors
 CC and cis-elements from regulatory sequences that mediate DC-specific
 CC expression; to determine the degree of maturity of DC and to block
 CC transcription factors, by providing binding sites in DC. (A) provide

CC DC-specific expression of nucleic acid under their control, allowing a
CC more specific regulation of the immune response and eliminating the long
CC and laborious purification of DC (since a complete leucocyte population
CC may be transformed), including transformation in vitro. This sequence
CC represents a DNA fragment of human fascin which is described in the
CC invention.

SO Sequence 1619 BP: 278 A; 466 C; 583 G; 292 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1619;

Best Local Similarity 100.0%; Pred. No. 41;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcacgcacgacatt 14

Db 787 gcacgcacgacatt 800

RESULT 18

AAH48552

ID AAH48552 standard; DNA: 1934 BP.

AC AAH48552;

DT 20-SEP-2001 (first entry)

DE Human fascin DNA fragment SEQ ID 2.

KW Fascin: regulatory sequence: human; dendritic cell; antiviral; tumor;
KW antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
KW immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
KW Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
KW autoimmune disease; transplant rejection; ds.

OS Homo sapiens.

PN WO200151631-A2.

PD 19-JUL-2001.

PF 12-JAN-2001; 2001WO-EP00362.

PR 13-JAN-2000; 2000DE-1001169.

PR 02-MAR-2000; 2000DE-1010188.

PA (RESK/) RESKE-KUNZ A.

PA (ROSS/) ROSS X.

PA (ROSS/) ROSS R.

PA (BROS/) BROS M.

PI Reske-Kunz A, Ross X, Ross R, Bros M;

DR WPI: 2001-451858/48.

PT New regulatory sequences from the fascin gene, useful for providing
PT dendritic cell-specific expression of e.g. antigens, e.g. for
PT vaccination against tumors and infections

XX Claim 1c; Page 85; 117pp; German.

CC This invention describes novel regulatory sequences (A) derived from
CC human fascin that provide specific expression in dendritic cells (DC) and
CC which have antiviral, antibacterial, antifungal, antiparasitic,
CC anti-allergic, neurological, immunomodulatory and apoptotic activity.

CC (A) are used to regulate expression of antigens, immunoregulators,
CC antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors
CC and host cells that contain (A) are useful: (i) in vaccines against
CC viruses, bacteria, fungi, parasites, tumors, allergens and plaques in
CC Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of
CC tumors, allergies, infections, autoimmune diseases and transplant
CC rejection. They can also be provide specific expression of antigens and
CC immunoregulators in DC; for isolation and identification of cell factors

CC and cis-elements from regulatory sequences that mediate DC-specific
CC expression; to determine the degree of maturity of DC and to block
CC transcription factors, by providing binding sites in DC. (A) provide
CC DC-specific expression of nucleic acid under their control, allowing a
CC more specific regulation of the immune response and eliminating the long
CC and laborious purification of DC (since a complete leucocyte population
CC may be transformed), including transformation in vitro. This sequence
CC represents a DNA fragment of human fascin which is described in the
CC invention.

SO Sequence 1934 BP: 344 A; 559 C; 647 G; 384 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1934;

Best Local Similarity 100.0%; Pred. No. 41;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcacgcacgacatt 14

Db 1102 gcacgcacgacatt 1115

RESULT 19

AAH48551

ID AAH48551 standard; DNA: 3069 BP.

AC AAH48551;

DT 20-SEP-2001 (first entry)

DE Human fascin DNA fragment SEQ ID 1.

KW Fascin: regulatory sequence: human; dendritic cell; antiviral; tumor;
KW antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
KW immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
KW Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
KW autoimmune disease; transplant rejection; ds.

OS Homo sapiens.

PN WO200151631-A2.

PD 19-JUL-2001.

PF 12-JAN-2001; 2001WO-EP00362.

PR 13-JAN-2000; 2000DE-1001169.

PR 02-MAR-2000; 2000DE-1010188.

PA (RESK/) RESKE-KUNZ A.

PA (ROSS/) ROSS X.

PA (ROSS/) ROSS R.

PA (BROS/) BROS M.

PI Reske-Kunz A, Ross X, Ross R, Bros M;

DR WPI: 2001-451858/48.

PT New regulatory sequences from the fascin gene, useful for providing
PT dendritic cell-specific expression of e.g. antigens, e.g. for
PT vaccination against tumors and infections

XX Claim 1a; Page 84-85; 117pp; German.

CC This invention describes novel regulatory sequences (A) derived from
CC human fascin that provide specific expression in dendritic cells (DC) and
CC which have antiviral, antibacterial, antifungal, antiparasitic,
CC anti-allergic, neurological, immunomodulatory and apoptotic activity.
CC (A) are used to regulate expression of antigens, immunoregulators,
CC antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors
CC and host cells that contain (A) are useful: (i) in vaccines against
CC viruses, bacteria, fungi, parasites, tumors, allergens and plaques in
CC Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of

CC tumors, allergies, infections, autoimmune diseases and transplant
 CC rejection. They can also be provide specific expression of antigens and
 CC immunoregulators in DC; for isolation and identification of cell factors
 CC and cis-elements from regulatory sequences that mediate DC-specific
 CC expression; to determine the degree of maturity of DC and to block
 CC transcription factors, by providing binding sites in DC. (A) provide
 CC DC-specific expression of nucleic acid under their control, allowing a
 CC more specific regulation of the immune response and eliminating the long
 CC and laborious purification of DC (since a complete leucocyte population
 CC may be transformed), including transformation in vitro. This sequence
 CC represents a DNA fragment of human fascin which is described in the
 CC invention.
 CC
 SQ Sequence 3069 BP; 607 A; 867 C; 993 G; 602 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 3069;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 gcacgcgcagcatt 14
 |||||
 DB 2237 gcacgcgcagcatt 2250

RESULT 20
 AAH48581
 ID AAH48581 standard; DNA; 4100 BP.

AAH48581;
 20-SEP-2001 (first entry)

DE Human fascin DNA fragment SEQ ID 33.

XX Fascin; regulatory sequence; human; dendritic cell; antiviral; tumor;
 KW antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
 KW immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
 KW Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
 KW autoimmune disease; transplant rejection; ds.

OS Homo sapiens.

PN WO200151631-A2.

PD 19-JUL-2001.

PF 12-JAN-2001; 2001WO-EP00362.

PR 13-JAN-2000; 2000DE-1001169.

PR 02-MAR-2000; 2000DE-1010188.

PA (RESK/) RESKE-KUNZ A.

PA (ROSS/) ROSS X.

PA (ROSS/) ROSS R.

PA (BROS/) BROS M.

PI Reske-Kunz A, Ross X, Ross R, Bros M;

XX WPI; 2001-451858/48.

DR New regulatory sequences from the fascin gene, useful for providing
 XX dendritic cell-specific expression of e.g. antigens, e.g. for
 PT vaccination against tumors and infections

XX Example 1; Page 102-103; 117pp; German.

CC This invention describes novel regulatory sequences (A) derived from
 CC human fascin that provide specific expression in dendritic cells (DC) and
 CC which have antiviral, antibacterial, antifungal, antiparasitic,
 CC anti-allergic, neurological, immunomodulatory and apoptotic activity.
 CC (A) are used to regulate expression of antigens, immunoregulators,
 CC antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors

CC and host cells that contain (A) are useful: (i) in vaccines against
 CC viruses, bacteria, fungi, parasites, tumors, allergens and plaques in
 CC Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of
 CC tumors, allergies, infections, autoimmune diseases and transplant
 CC rejection. They can also be provide specific expression of antigens and
 CC immunoregulators in DC; for isolation and identification of cell factors
 CC and cis-elements from regulatory sequences that mediate DC-specific
 CC expression; to determine the degree of maturity of DC and to block
 CC transcription factors, by providing binding sites in DC. (A) provide
 CC DC-specific expression of nucleic acid under their control, allowing a
 CC more specific regulation of the immune response and eliminating the long
 CC and laborious purification of DC (since a complete leucocyte population
 CC may be transformed), including transformation in vitro. This sequence
 CC represents a DNA fragment of human fascin which is described in the
 CC invention.
 CC
 SQ Sequence 4100 BP; 769 A; 1251 C; 1339 G; 741 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 4100;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcacgcgcagcatt 14
 |||||
 DB 2237 gcacgcgcagcatt 2250

RESULT 21

AAZ49334
 ID AAZ49334 standard; cDNA; 4189 BP.

XX AAZ49334;

AC 14-MAR-2000 (first entry)

DE Murine multidrug resistance-1 (MDR-1) cDNA.

XX Multidrug resistance; MDR-1; P-glycoprotein;
 KW transmembrane efflux pump; hematopoietic stem cell; transduction;
 KW bone marrow transplantation; chemotherapy; radiation therapy; cancer;
 KW gene therapy; gene replacement; genetic defect; thalassemia;
 KW Gaucher's disease; sickle cell anaemia; leukaemia; ex vivo expansion;
 KW cytokine; ds.

XX Mus sp.

XX Key Location/Qualifiers

XX CDS 1..3831

XX FT /*tag= a /product= "Murine MDR-1 protein"

PN WO9961589-A2.

XX 02-DEC-1999.

PF 27-MAY-1999; 99WO-US11825.

PR 28-MAY-1998; 98US-0086988.

PA (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.

PA Sorrentino B, Bunting K;

XX WPI; 2000-072615/06.

DR P-PSDB; AAY58188.

XX Ex vivo expansion of hematopoietic stem cells transduced with a
 PT sequence encoding human multidrug resistance-1, used for bone marrow
 PT transplantation

XX Disclosure; Page 90-93; 113pp; English.

CC This sequence represents cDNA encoding murine multidrug
CC resistance protein MDR-1. MDR-1 is a transmembrane
CC efflux pump, responsible for the export of drugs from cells.
CC particularly cancer cells. The invention relates to transducing
CC haematopoietic stem cells with nucleic acid encoding an MDR protein
CC and culturing the modified cells. The modified haematopoietic stem
CC cells are useful in bone marrow transplantation (to reconstitute
CC haematopoietic systems in patients who have undergone chemotherapy or
CC radiation therapy) and in *ex vivo* gene therapy of genetic defects in
CC cells derived from haematopoietic stem cells, e.g., thalassemia,
CC Gaucher's disease, sickle cell anaemia or leukaemia. The modified
CC cells can also be used to identify factors involved in regulating
CC proliferation and differentiation in haematopoietic stem cells.
CC Haematopoietic stem cells that express MDR-1 will be protected against
CC chemotherapeutic agents, so can be engrafted while the patient is
CC undergoing chemotherapy. Expansion of (rare) haematopoietic stem cells
CC provides sufficient cells to permit standard biochemical analysis.
CC Overexpression of MDR-1 allows cytokine-driven expansion of
CC haematopoietic stem cells by at least 10-fold compared with a maximum
CC of 4-fold in known procedures.
XX
S0 Sequence 4189 BP; 1204 A; 875 C; 1028 G; 1082 T; 0 other;

Query Match	56.0%	Score 14	DB 21	length 4189
Best Local Similarity	100.0%	Pred. No.	40	
Matches 14	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	6 gccagcattcgat	19		
Db	880 gccagcattcgat	893		

RESULT	22
AAQ38950	
ID	AAQ38950 standard; DNA; 4313 BP.
XX	
AC	AAQ38950;
XX	
DT	28-JUL-1993 (first entry)
XX	
DE	Mouse multidrug resistance sequence.
XX	
XX	mdr gene; Lambda DR11 clone; ss.
XX	
OS	Mus musculus.
XX	
FH	Key
FT	Location/Qualifiers
FT	CDS
FT	110..3940
FT	/tag= a
FT	/phenotype= multidrug_resistance
XX	
XX	US5198344-A.
XX	
PD	30-MAR-1993.
XX	
PF	15-JUL-1986; 86US-0885951.
XX	
XX	15-JUL-1986; 86US-0885951.
PR	06-FEB-1991; 91US-0652311.
XX	
PA	(MAST) MASSACHUSETTS INST TECHNOLOGY.
XX	
PI	Groop JM, Gros P, Housman DE;
XX	
DR	WPI: 1993-126077/15.
DR	P-PSDB; AAR35199.
XX	
PT	DNA sequence which confers multi-drug resistance on sensitive
PT	mammalian cells - used to preserve bone marrow cells during
XX	chemotherapy to prevent infection
PS	Claim 1; Fig 8; 22pp; English.

XX A cDNA library was constructed from mRNA which had been isolated
CC from a drug-sensitive mouse cell-line. Two mouse cDNA molecules
CC complementary to the mRNA species encoded by 2 related but distinct
CC mdr genes were isolated and cloned. One of the cDNA clones (lambda
CC DR11) is a full-length cDNA clone for one member of the mdr gene
CC family. When incorporated into prokaryotic expression vector pPREX4
CC (which allows high levels of transcription of the cDNA when
CC introduced into mammalian cells), the clone was shown to confer the
CC multidrug resistance phenotype upon transfection into drug-sensitive
CC mammalian cells.
XX
SQ Sequence 4313 BP: 1241 A; 904 C; 1058 G; 1110 T; 0 other;

Query Match	56.0%;	Score 14;	DB 14;	Length 4313;
Best Local Similarity	100.0%;	Pred. No. 40;		
Matches 14;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	6 gccacacattcgat	19		
Db	989 gccacacattcgat	1002		

RESULT	23
AAV35235/C	
ID	AAV35235 standard; DNA; 5319 BP.
XX	
AC	AAV35235;
XX	
DT	28-SEP-1998 (first entry)
XX	
DE	R. prowazekii S-layer protein genomic DNA.
XX	
KW	Surface layer protein; S-layer; vaccine; typhus; spotted fever;
XX	
OS	infection; diagnosis; disease; ds.
XX	
OS	Rickettsia prowazekii.
XX	
PH	Key
FT	-35_signal
FT	340..345
FT	/*tag= a
FT	363..368
FT	-10_signal
FT	/*tag= b
FT	379..386
FT	/*tag= c
FT	391..5229
FT	/*tag= d
FT	/product= S-layer protein
FT	5270..5306
FT	/*tag= e
XX	
PN	US5783441-A.
XX	
XX	
PD	21-JUL-1998.
XX	
PE	20-DEC-1993; 93US-0169927.
XX	
PR	20-DEC-1993; 93US-0169927.
PR	09-AUG-1991; 91US-0742128.
XX	
PA	(USNA) US SEC OF NAVY.
XX	
PI	Carl M, Ching W, Dasch GA, Dobson ME;
XX	
DR	WPI: 1998-427031/36.
DR	P-PSDB: AAW65088.
XX	
PT	Recombinant DNA encoding Rickettsia surface layer proteins - useful
PT	for diagnosing typhus and spotted fever and for preparing vaccines
PT	against them
XX	
PS	Claim 1: Column 11-24; 20pp; English.

XX This sequence encodes the Surface layer (S-layer) protein from
 CC R. prowazekii strain Breinl. This sequence is useful for vaccination
 CC against typhus and spotted fever rickettsial infection or for diagnosing
 CC diseases caused by these bacteria. The surface layer protein antigens can
 CC be produced recombinantly in large quantities.
 CC
 SQ Sequence 5319 BP; 1815 A; 766 C; 983 G; 1755 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 5319;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 catcgccagcattt 15
 |||||||||||||
 Db 2221 CATGCCAGCATTT 2208

RESULT 24
 ID AAH48621 standard; DNA: 13055 BP.
 XX
 AC AAH48621;
 XX
 DT 20-SEP-2001 (first entry)
 XX
 DE Human fascin DNA fragment #1.
 XX
 KW Fascin; regulatory sequence; human; dendritic cell; antiviral; tumor;
 KW antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
 KW immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
 KW Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
 KW autoimmune disease; transplant rejection; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200151631-A2.
 XX
 PD 19-JUL-2001.
 XX
 PF 12-JAN-2001; 2001WO-EP00362.
 XX
 PR 13-JAN-2000; 2000DE-1001169.
 XX
 PR 02-MAR-2000; 2000DE-1010188.
 XX
 PA (RESK/) RESKE-KUNZ A.
 PA (ROSS/) ROSS X.
 PA (ROSS/) ROSS R.
 PA (BROS/) BROS M.
 XX
 PI Reske-Kunz A, Ross X, Ross R, Bros M;
 XX
 DR WPI; 2001-451858/48.
 XX
 PT New regulatory sequences from the fascin gene, useful for providing
 PT dendritic cell-specific expression of e.g. antigens, e.g. for
 PT vaccination against tumors and infections
 XX
 PS Example 1; Fig 2; 117pp; German.
 XX
 CC This invention describes novel regulatory sequences (A) derived from
 CC human fascin that provide specific expression in dendritic cells (DC) and
 CC which have antiviral, antibacterial, antifungal, antiparasitic,
 CC anti-allergic, neurological, immunomodulatory and apoptotic activity.
 CC (A) are used to regulate expression of antigens, immunoregulators,
 CC antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors
 CC and host cells that contain (A) are useful: (i) in vaccines against
 CC viruses, bacteria, fungi, parasites, tumors, allergens and plaques in
 CC Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of
 CC tumors, allergies, infections, autoimmune diseases and transplant
 CC rejection. They can also be provide specific expression of antigens and
 CC immunoregulators in DC; for isolation and identification of cell factors

CC and cis-elements from regulatory sequences that mediate DC-specific
 CC expression; to determine the degree of maturity of DC and to block
 CC transcription factors, by providing binding sites in DC. (A) provide
 CC DC-specific expression of nucleic acid under their control, allowing a
 CC more specific regulation of the immune response and eliminating the long
 CC and laborious purification of DC (since a complete leucocyte population
 CC may be transformed), including transformation in vitro. This sequence
 CC represents a DNA fragment of human fascin which is described in the
 CC invention.
 CC
 SQ Sequence 13055 BP; 2372 A; 3903 C; 4097 G; 2683 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 13055;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcacgcagcattt 14
 |||||||||||||
 Db 2237 gcacgcagcattt 2250

RESULT 25
 ID AAH48620 standard; DNA: 16951 BP.
 XX
 AC AAH48620;
 XX
 DT 20-SEP-2001 (first entry)
 XX
 DE Human fascin DNA fragment SEQ ID 72.
 XX
 KW Fascin; regulatory sequence; human; dendritic cell; antiviral; tumor;
 KW antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
 KW immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
 KW Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
 KW autoimmune disease; transplant rejection; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200151631-A2.
 XX
 PD 19-JUL-2001.
 XX
 PF 12-JAN-2001; 2001WO-EP00362.
 XX
 PR 13-JAN-2000; 2000DE-1001169.
 XX
 PR 02-MAR-2000; 2000DE-1010188.
 XX
 PA (RESK/) RESKE-KUNZ A.
 PA (ROSS/) ROSS X.
 PA (ROSS/) ROSS R.
 PA (BROS/) BROS M.
 XX
 PI Reske-Kunz A, Ross X, Ross R, Bros M;
 XX
 DR WPI; 2001-451858/48.
 XX
 PT New regulatory sequences from the fascin gene, useful for providing
 PT dendritic cell-specific expression of e.g. antigens, e.g. for
 PT vaccination against tumors and infections
 XX
 PS Claim 1a; Page 112-117; 117pp; German.
 XX
 CC This invention describes novel regulatory sequences (A) derived from
 CC human fascin that provide specific expression in dendritic cells (DC) and
 CC which have antiviral, antibacterial, antifungal, antiparasitic,
 CC anti-allergic, neurological, immunomodulatory and apoptotic activity.
 CC (A) are used to regulate expression of antigens, immunoregulators,
 CC antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors
 CC and host cells that contain (A) are useful: (i) in vaccines against
 CC viruses, bacteria, fungi, parasites, tumors, allergens and plaques in
 CC Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of

CC oligonucleotides and compositions (I) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'universal' or alternative base.
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antiallergic, hypotensive and cytostatic activities.
 CC The antisense oligonucleotides and (I) can be used to down-regulate the
 CC expression and/or activity of target polypeptides associated with the
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokines and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, defensins, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impaired respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.

XX Sequence 143068 BP; 41194 A; 30122 C; 32403 G; 39349 T; 0 other;

XX Query Match 56.0%; Score 14; DB 21; Length 143068;
 XX Best Local Similarity 100.0%; Pred. No. 35;
 XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cagcattcgatcc 21
 DB 89829 CAGCATTTCGATCC 89816

RESULT 28
 AAF21272/c
 ID AAF21272 standard; DNA: 143068 BP.

XX AAF21272;

DT 14-MAR-2001 (first entry)

DE Human low adenosine antisense oligonucleotide related sequence #2839.

XX Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
 KW human; airway disorder; bronchoconstriction; lung inflammation;
 KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
 KW immunosuppressive; antiallergic; hypotensive; cytostatic;
 KW respiratory obstruction; pulmonary obstruction; impaired respiration;
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.

XX Homo sapiens.

XX WO200062736-A2.

XX 26-OCT-2000.

XX 24-MAR-2000; 2000WO-US08020.

XX 06-APR-1999; 99US-0127958.

XX (UYEC-) UNTV EAST CAROLINA.

PA (NYCE/) NYCE J W.
 XX NYCE JW;
 PI NYCE JW;
 XX WPI: 2000-679539/66.

PT Low adenosine (A) content antisense oligonucleotides which do not
 PT trigger adenosine receptors during metabolism, useful e.g. for treating
 PT cancers and respiratory obstructions -

PS Disclosure, Page 1186-1219; 1592pp; English.

CC The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (I) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'universal' or alternative base.
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antiallergic, hypotensive and cytostatic activities.
 CC The antisense oligonucleotides and (I) can be used to down-regulate the
 CC expression and/or activity of target polypeptides associated with the
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, defensins, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impaired respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.

XX Sequence 143068 BP; 41194 A; 30122 C; 32403 G; 39349 T; 0 other;

XX Query Match 56.0%; Score 14; DB 21; Length 143068;
 XX Best Local Similarity 100.0%; Pred. No. 35;
 XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cagcattcgatcc 21
 DB 89829 CAGCATTTCGATCC 89816

RESULT 29
 AAA34983/c
 ID AAA34983 standard; DNA: 143068 BP.

XX AAA34983;

DT 28-JUL-2000 (first entry)

DE Human adenosine receptor related polynucleotide SEQ ID NO:2672.

XX Human; adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antiallergic; cytostatic; analgesic; impaired airway;
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

OS Homo sapiens.
 XX
 PN W0200009525-A2.
 XX
 PD 24-FEB-2000.
 XX
 PF 03-AUG-1999; 99WO-US17712.
 XX
 PR 03-AUG-1998; 98US-0095212.
 XX
 PA (UYEC-) UNIV EAST CAROLINA.
 XX
 PI Nyce JW;
 XX
 DR WPI: 2000-205971/18.
 XX
 PT New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
 PT cancers -
 XX
 PS Disclosure: Page 851-882; 1343pp; English.
 XX
 CC The present invention describes a new composition comprising an
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
 CC inflammation. The ON can have antiinflammatory, antiallergic,
 CC antiasthmatic, cytosstatic and analgesic activities. The compositions are
 CC useful for the treatment of diseases associated with inflammation,
 CC impaired airways, including lung disease and diseases whose secondary
 CC effects afflict the lungs of a subject. They can be used for treating
 CC e.g. ischemic conditions, pulmonary vasoconstriction, allergies,
 CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC carcinomas, and cancers which may metastasize to the lungs, including
 CC breast and prostate cancer. The reduction of the adenosine content of
 CC the ONs reduces side effects. The A-containing ONs break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680
 CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
 CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.
 CC
 XX
 SQ Sequence 143068 BP; 41194 A; 30122 C; 32402 G; 39350 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 143068;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cagcattcgatcc 21
 |||
 DB 89829 CAGCATTTCGATCC 89816

RESULT 30
 ID AAA35150/c
 AC AAA35150 standard; DNA; 143068 BP.
 XX
 XX AAA35150;
 XX
 DT 28-JUL-2000 (first entry)
 XX
 DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:24.
 XX
 XX Human: adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphothioate; impaired respiration; inflammation; allergy;

KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW anti-allergic; antiasthmatic; cytosstatic; analgesic; impeded airway;
 KW lung disease; ischemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukemia; lymphoma; carcinoma; metastasis; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200009525-A2.
 XX
 PD 24-FEB-2000.
 XX
 PF 03-AUG-1999; 99WO-US17712.
 XX
 PR 03-AUG-1998; 98US-0095212.
 XX
 PA (UYEC-) UNIV EAST CAROLINA.
 XX
 PI Nyce JW;
 XX
 DR WPI: 2000-205971/18.
 XX
 PT New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
 PT cancers -
 XX
 PS Disclosure: Page 1106-1138; 1343pp; English.
 XX
 CC The present invention describes a new composition comprising an
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
 CC inflammation. The ON can have antiinflammatory, antiallergic,
 CC antiasthmatic, cytosstatic and analgesic activities. The compositions are
 CC useful for the treatment of diseases associated with inflammation,
 CC impaired airways, including lung disease and diseases whose secondary
 CC effects afflict the lungs of a subject. They can be used for treating
 CC e.g. ischemic conditions, pulmonary vasoconstriction, allergies,
 CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC carcinomas, and cancers which may metastasize to the lungs, including
 CC breast and prostate cancer. The reduction of the adenosine content of
 CC the ONs reduces side effects. The A-containing ONs break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680
 CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
 CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.
 CC
 XX
 SQ Sequence 143068 BP; 41194 A; 30126 C; 32402 G; 39346 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 143068;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cagcattcgatcc 21
 |||
 DB 89829 CAGCATTTCGATCC 89816

RESULT 31
 ID AAA35151/c
 AC AAA35151 standard; DNA; 149412 BP.
 XX
 XX AAA35151;
 AC

XX 28-JUL-2000 (first entry)
 DT
 XX Human adenosine receptor related polynucleotide 2nd SEQ ID NO:25.
 DE
 XX
 KW Human: adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antiasthmatic; cytosolic; analgesic; impaired airway;
 KW lung disease; ischemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukemia; lymphoma; carcinoma; metastasis; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200009525-A2.
 XX
 PD 24-FEB-2000.
 XX
 PF 03-AUG-1999; 99WO-US17712.
 XX
 PR 03-AUG-1998; 98US-0095212.
 XX
 PA (UYEC-) UNIV EAST CAROLINA.
 XX
 PI Nyce JW;
 XX
 DR WPI: 2000-205971/18.
 XX
 PT New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
 PT cancers -
 XX
 PS Disclosure; Page 1138-1171; 1343pp; English.
 XX
 CC The present invention describes a new composition comprising an
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
 CC inflammation. The ON can have antiinflammatory, antiallergic,
 CC antiasthmatic, cytosolic and analgesic activities. The compositions are
 CC useful for the treatment of diseases associated with inflammation,
 CC impaired airways, including lung disease and diseases with secondary
 CC effects afflict the lungs of a subject. They can be used for treating
 CC e.g. ischemic conditions, pulmonary vasoconstriction, allergies,
 CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC carcinomas, and cancers which may metastasize to the lungs, including
 CC breast and prostate cancer. The reduction of the adenosine content of
 CC the ONs reduces side effects. The A-containing ONs break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing the
 CC bronchoconstriction and inflammation. AAA3213 to AAA35312 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 CC differ from the previously named sequences. SEQ ID NO:11 to 1660
 CC (AAA3213 to AAA3992) are specifically claimed ONs from the present
 CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.
 XX
 XX Sequence 149412 BP; 43049 A; 31388 C; 33852 G; 41123 T; 0 other;
 SQ

Query Match 56.0%; Score 14; DB 21; Length 149412;
 Best Local Similarity 100.0%; Pired. No. 35;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

07 8 cagcattcagatcc 21
 |||||||||
 Db 96173 CAGCATTTGCATCC 96160

RESULT 32
 AAF21273/C
 ID AAF21273 standard; DNA; 152740 BP.
 XX
 XX AAF21273;
 AC
 XX
 DT 14-MAR-2001 (first entry)
 XX
 DE Human low adenosine antisense oligonucleotide related sequence #2840.
 XX
 KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
 KW human; airway disorder; bronchoconstriction; lung inflammation;
 KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
 KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytosolic;
 KW respiratory obstruction; pulmonary obstruction; impaired respiration;
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200062736-A2.
 XX
 PD 26-OCT-2000.
 XX
 PF 24-MAR-2000; 2000WO-US08020.
 XX
 PR 06-APR-1999; 99US-0127958.
 XX
 PA (UYEC-) UNIV EAST CAROLINA.
 PA (NYCE/) NYCE J W.
 XX
 PI Nyce JW;
 XX
 DR WPI: 2000-679539/66.
 XX
 PT Low adenosine (A) content antisense oligonucleotides which do not
 PT trigger adenosine receptors during metabolism, useful e.g. for treating
 PT cancers and respiratory obstructions -
 XX
 PS Disclosure; Page 1219-1254; 1592pp; English.
 XX
 CC The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (I) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antiasthmatic, hypotensive and cytosolic activities.
 CC The antisense oligonucleotides and (I) can be used to down-regulate the
 CC expression and/or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, defensins, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impaired respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide

CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.

XX Sequence 152740 BP; 44169 A; 32023 C; 34549 G; 41999 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 152740;

Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 cagcattcgatcc 21
|||||

DB 96173 CAGCATTTCGATCC 96160

RESULT 33
AAV1150/C
ID AAV1150 standard; DNA; 22 BP.

XX AAV1150;

XX 14-JUL-1998 (first entry)

DE Oligonucleotide #3 for nucleic acid assay.

XX Nucleic acid assay; hybridising; probe; target; duplex; detection; ss.

XX Synthetic.

XX DE19633436-A1.

XX 26-FEB-1998.

XX 20-AUG-1996; 96DE-1033436.

XX 20-AUG-1996; 96DE-1033436.

XX (BOE) BOEHRINGER MANNHEIM GMBH.

XX Bergmann F, Herrmann R, Kobold U;

XX WPI; 1998-146248/14.

XX Nucleic acid assay - based on mass of probe-target duplex

XX Disclosure; Fig 1; 9pp; German.

XX AAV1148-V1151 are oligonucleotides used in a novel nucleic acid assay.

XX This assay involves hybridising a probe to a target sequence of the

XX nucleic acid to form a duplex and degrading the duplex, which is detected

XX a fragment of a defined length containing the duplex, which is detected

XX on the basis of its mass. This method enables different nucleic acids to

XX be detected simultaneously does not require the use of labelled probes.

XX Sequence 22 BP; 5 A; 3 C; 7 G; 7 T; 0 other;

OY 2 cagcagcattc 14
|||||

DB 18 CATGCCGATTC 6

RESULT 34
AAA76686
ID AAA76686 standard; cDNA; 51 BP.

XX AAA76686;

XX 16-NOV-2000 (first entry)

XX Human clone cg28389525 polymorphic site, SEQ ID NO:369.

DE Human: single nucleotide polymorphism; SNP;
XX detection; Identification; gene therapy; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
XX variation replace (26,C)
XX /*tag= a

FT WO200029623-A2.

XX 25-MAY-2000.

XX 17-NOV-1999; 99WO-US27293.

XX 17-NOV-1998; 98US-0109024.

XX 16-NOV-1999; 99US-0109024.

XX (CURA-) CURAGEN CORP.

XX Shimkels RA, Leach MD;

XX WPI; 2000-387826/33.

XX Human nucleic acids containing single nucleotide polymorphisms, useful

XX for treating a subject suffering, or at risk from a pathology due to

XX the presence of a sequence polymorphism -

XX Claim 1; Page 269; 543pp; English.

XX Sequences AAA76318-A77509 represent 1192 human nucleic acid sequences

XX which contain single nucleotide polymorphisms (SNPs). Sequences 1 to

XX 1112 (AAA76318-A77429) are consecutive pairs of nucleotides which

XX contain silent SNPs. Sequences 1113 to 1192 (AAA77430-A77509) are

XX consecutive pairs of nucleotides containing SNPs which result in changes

XX in the corresponding amino acid sequences (AAA11749-B11828). The SNPs in

XX sequences 1113 to 1128 (AAA77430-A77445) lead to conservative amino acid

XX changes, while those in sequences 1129 to 1166 (AAA77446-A77503) result

XX in non-conservative changes. The SNPs in sequences 1167 to 1192

XX relates to a method of detecting a polymorphic site in a nucleic acid and

XX a method of determining the relatedness of two nucleic acids. It also

XX encompasses peptides containing polymorphic sites, antibodies raised

XX against such peptides, and a method of detecting polymorphic

XX proteins/peptides using the antibodies. The nucleic acids are useful for

XX gene therapy of an individual having, suspected of having, or at risk of

XX developing a pathological condition due to the presence of a sequence

XX polymorphism. Such treatment would comprise administration of the

XX wild-type nucleic acid sequence. Antibodies raised against polymorphic

XX peptides can also be used in the treatment of such individuals.

XX Sequence 51 BP; 13 A; 11 C; 15 G; 12 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 51;

Best Local Similarity 100.0%; Pred. No. 17e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 cagcagcattcg 17
|||||

DB 11 cgcagcattcg 23

RESULT 35
AAA76687
ID AAA76687 standard; cDNA; 51 BP.

XX AAA76687;

XX 16-NOV-2000 (first entry)

```

XX DE Human clone cg28389525 polymorphic site, SEQ ID NO:370.
XX KW Human: single nucleotide polymorphism; SNP;
XX KW detection; identification; gene therapy; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT variation /tag=a
XX PN WO200029623-A2.
XX PD 25-MAY-2000.
XX PF 17-NOV-1999; 99WO-US27293.
XX PR 17-NOV-1998; 98US-0109024.
XX PR 16-NOV-1999; 99US-0109024.
XX PA (CURA-) CURAGEN CORP.
XX PI Shimkets RA, Leach MD;
XX DR WPI; 2000-387826/33.
XX PT Human nucleic acids containing single nucleotide polymorphisms, useful
XX PT for treating a subject suffering, or at risk from a pathology due to
XX PT the presence of a sequence polymorphism.
XX PS Claim 1; Page 270; 543pp: English.
XX CC Sequences AAAT76318-A77509 represent 1192 human nucleic acid sequences
XX CC which contain single nucleotide polymorphisms (SNPs). Sequences 1 to
XX CC 1112 (AAAT76318-A77429) are consecutive pairs of nucleotides which
XX CC contain silent SNPs. Sequences 1113 to 1192 (AAAT7430-A77509) are
XX CC consecutive pairs of nucleotides containing SNPs which result in changes
XX CC in the corresponding amino acid sequences (AAAT1749-B11828). The SNPs in
XX CC sequences 1113 to 1128 (AAAT7430-A77445) lead to conservative amino acid
XX CC changes, while those in sequences 1129 to 1186 (AAAT7446-A77503) result
XX CC in non-conservative changes. The SNPs in sequences 1187 to 1192
XX CC (AAAT7504-A77509) generate frameshift mutations. The invention also
XX CC relates to a method of detecting a polymorphic site in a nucleic acid and
XX CC a method of determining the relatedness of two nucleic acids. It also
XX CC encompasses peptides containing polymorphic sites, antibodies raised
XX CC against such peptides, and a method of detecting polymorphic
XX CC proteins/peptides using the antibodies. The nucleic acids are useful for
XX CC gene therapy of an individual having, suspected of having, or at risk of
XX CC developing a pathological condition due to the presence of a sequence
XX CC polymorphism. Such treatment would comprise administration of the
XX CC wild-type nucleic acid sequence. Antibodies raised against polymorphic
XX CC peptides can also be used in the treatment of such individuals.
XX SQ Sequence 51 BP; 13 A; 12 C; 15 G; 11 T; 0 other:

```

```

Query Match 52.0%; Score 13; DB 21; Length 51;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 5 cgcagcatttcg 17
   |||||||
Db 11 cgcagcatttcg 23

```

```

RESULT 36
AAZ19606
ID AAZ19606 standard; RNA; 61 BP.
AC AAZ19606;
XX
DT 08-NOV-1999 (first entry)

```

```

XX DE Complement system protein C5 RNA binding ligand 13.
XX KW Complement system protein; C1q; C5; C3; RNA ligand; nootropic; stroke;
XX KW neuroprotective; anti-asthmatic; cerebroprotective; immunosuppressive;
XX KW diagnostic; prevention; treatment; complement protein-related disease;
XX KW Alzheimer's disease; renal disease; transplant rejection; asthma; ss.
XX OS Synthetic.
XX FH Synthesis
XX FT WO9941271-A1.
XX PN 19-AUG-1999.
XX PD 05-FEB-1999; 99WO-US02597.
XX PF 29-SEP-1998; 98US-0163025.
XX PR 12-FEB-1998; 98US-0023228.
XX PA (NEXS-) NEXSTAR PHARM INC.
XX PI Biesecker G, Gold L;
XX DR WPI; 1999-527357/44.
XX PT New Nucleic Acid Ligand to complement protein C5, useful in the
XX PT treatment of Alzheimer's disease, renal diseases, transplant
XX PT rejection, stroke and asthma
XX PS Example 4; Page 42; 120pp: English.
XX CC This invention describes novel purified and isolated non-naturally
XX CC occurring RNA ligands (I) to complement system proteins C5, C1q and C3.
XX CC The products of the invention have nootropic, neuroprotective,
XX CC anti-asthmatic, cerebroprotective and immunosuppressive activity. The
XX CC new ligands are useful as diagnostic agents, and pharmaceutical agents
XX CC for prevention and treatment of complement protein-related diseases,
XX CC including Alzheimer's disease, renal diseases, transplant rejection,
XX CC stroke and asthma.
XX SQ Sequence 61 BP; 11 A; 18 C; 21 G; 11 U; 0 other:

```

```

Query Match 52.0%; Score 13; DB 20; Length 61;
Best Local Similarity 69.2%; Pred. No. 1.7e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 8 cagcattcagtc 20
   |||:::|
Db 23 cagcauuucgac 35

```

```

RESULT 37
AAZ14744/c
ID AAZ14744 standard; cDNA; 300 BP.
XX
AC AAZ14744;
XX
DT 12-OCT-1999 (first entry)
XX

```

```

Human gene expression product cDNA sequence SEQ ID NO:2213.
DE
XX
XX KW Human: gene expression product; diagnosis; therapy; probe;
XX KW detection; mapping; tissue typing; profiling; forensic; cancer;
XX KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX OS Homo sapiens.
XX PN WO9938972-A2.
XX PD 05-AUG-1999.
XX PF 28-JAN-1999; 99WO-US01619.

```

XX 03-APR-1998; 98US-0080666.
 PR 28-JAN-1998; 98US-0072910.
 PR 24-FEB-1998; 98US-0075954.
 PR 31-MAR-1998; 98US-0080114.
 PR 03-APR-1998; 98US-0080515.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE) HYSEQ INC.
 XX
 PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
 PI Stache-Crain B, Suduth-Klinger J, Williams LT;
 XX
 DR WPI; 1999-494092/41.

PT Novel human genes and their expression products which are
 PT differentially expressed in different cell types

PS Claim 1: Page 1159; 2479pp; English.

XX The present invention describes a library of human polynucleotides
 CC comprising the sequences given in AA212532 to AA217779. Also described is
 CC a method of detecting differentially expressed genes correlated with the
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample from a cell
 CC suspected of being cancerous, where the gene product is encoded by one
 CC of the 5248 polynucleotide sequences given in AA212532 to AA217779. The
 CC polynucleotides can be used as a source of primers and probes, which can
 CC be used for a variety of purpose, e.g. detection of expression levels,
 CC mapping, tissue typing or profiling, forensics, genetic analysis and
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnostics (which may be used to determine function of an
 CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists.

XX Sequence 300 BP: 83 A; 62 C; 70 G; 85 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 300;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgcacagcat 13
 |||
 DB 174 GCATGCCACGACAT 162

RESULT 38
 AAA31174/c
 ID AAA31174 standard; DNA: 322 BP.

XX AAA31174;

DT 05-JUL-2000 (first entry)

XX Plant microsatellite marker #135.

XX Plant microsatellite sequence; core repeat sequence; detection: probe;

KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;

XX variety identification; genetic variability evaluation; primer; ss.

OS EucaIyplus grandis.

PN W0967421-A1.

XX 29-DEC-1999.

XX 25-JUN-1999; 99WO-NZ00092.

XX 25-JUN-1998; 98US-0105307.

PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.

PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX Havukkala JU, Bloksberg LN, Glenn M;

DR WPI; 2000-116958/10.

PT New plant microsatellite markers and associated flanking species for

PT the detection of polymorphic genetic markers -

PS Claim 1: Page 116; 392pp; English.

XX Sequences AAA31040-A32093 represent novel plant microsatellite sequences
 CC and associated flanking species. The sequences comprise a central core
 CC repeat sequence, especially selected from the sequences AAA32094-A32096
 CC with left and right flanking sequences. The polynucleotide sequences
 CC can be used in the detection of DNA polymorphisms, in genome mapping,
 CC in physical mapping, in positional cloning of genes, in variety
 CC identification and in evaluation of genetic variability within and
 CC between plant tissues, populations, cultivars, species and species
 CC groups. They may also be used to design hybridization probes for
 CC oligonucleotide fingerprinting and library screening and to design
 CC primers for microsatellite-primed PCR. Microsatellite markers are
 CC useful to locate specific economically useful genes in plant genomes.

XX Sequence 322 BP: 43 A; 101 C; 88 G; 89 T; 1 other;

Query Match 52.0%; Score 13; DB 21; Length 322;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atcgccagcat 15
 |||
 DB 21 ATCGCCAGCATTT 9

RESULT 39

AA114896

ID AA114896 standard; DNA: 435 BP.

XX AA114896;

DT 12-OCT-2001 (first entry)

DE Probe #4829 for gene expression analysis in human cervical cell sample.

KW Probe; human; microarray; gene expression; cervical epithelial cell;

KW cervical cancer; ss.

XX Homo sapiens.

PN W0200157278-A2.

XX 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00670.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0633366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID No 4829; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 435 BP; 88 A; 115 C; 98 G; 134 T; 0 other;

Query Match 52.0%; Score 13; DB 22; Length 435;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 8 cagcattcgatc 20
| | | | | | | | | |
Db 12 cagcattcgatc 24

RESULT 40
AA136249
ID AA136249 standard; DNA; 435 BP.
XX
AC AA136249;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #4935 used to measure gene expression in human placenta sample.
XX
KW Probe: microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-48897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX

PS Claim 25; SEQ ID No 4935; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENPs).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 435 BP; 88 A; 115 C; 98 G; 134 T; 0 other;

Query Match 52.0%; Score 13; DB 22; Length 435;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 8 cagcattcgatc 20
| | | | | | | | | |
Db 12 cagcattcgatc 24

RESULT 41
AA104678
ID AA104678 standard; DNA; 435 BP.
XX
AC AA104678;
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #4669 used to measure gene expression in human breast sample.
XX
KW Probe: human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
PN WO200157270-A2.
XX
PD 09-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US00661.
XX

PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-476286/51.
XX
PT Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast -
XX
PS Claim 25; SEQ ID No 4669; 322pp; English.
XX

XX The present invention relates to novel single exon nucleic acid probes.
XX The present sequence is one such probe. The probes are useful for
XX measuring human gene expression in a human breast sample, where the probe
XX hybridises at high stringency to a nucleic acid expressed in the human
XX breast. The probes are useful for predicting, diagnosing, grading,
XX staging, monitoring and prognosing diseases of the human breast,
XX particularly those diseases with polygenic aetiology. The diseases
XX include: breast cancer, disorders of development, inflammatory diseases
XX of the breast, fibrocystic changes, proliferative breast disease and
XX non carcinoma tumours.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 435 BP; 88 A; 115 C; 98 G; 134 T; 0 other;

Query Match 52.0%; Score 13; DB 22; Length 435;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 cagcattcgcac 20
| | | | | | | | | |
DB 12 cagcattcgcac 24

RESULT 42

AAT27962
ID AAT27962 standard; DNA: 447 BP.

AC AAT27962;
XX
DT 11-MAR-1997 (first entry)

DE Hepatitis C virus type 10a isolate NN98 bases 478-925.

KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;
PCR; primer; probe; antibody; infection; ss.

OS Hepatitis C virus.

PN WO9613590-A2.

PD 09-MAY-1996.

PF 23-OCT-1995; 95WO-EP04155.

PR 28-JUN-1995; 95EP-0870076.

PR 21-OCT-1994; 94EP-0870166.

XX (INNO-) INNOGENETICS NV.

PA Maertens G, Stuyver L;

PI WPI: 1996-251460/25.

DR P-PSDB; AAR96551.

PT Hepatitis C virus poly:nucleic acid unique to unidentified sub:type
PT - used to develop probes and primers for new sub:types and vaccines
to prevent and treat infection

PS Claim 6; Fig 3: 150pp; English.

XX The sequences AAT27937-127989 represent novel sequences isolated from
CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
CC 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-l, 2k, 2l,
CC 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
CC untranslated region (UR), the Core/E1, NS4 or NS5B regions of the
CC genome. This sequence represents nucleotides 478-925 from the HCV type
CC 10a isolate NN98.
CC The new HCV types were isolated from patients with chronic HCV from the
CC Benelux countries, France, Cameroon and Vietnam, because of their
CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
CC amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions were
CC sequenced either directly or partially and used to classify the new
CC viruses into (sub)types based on comparison with known sequences.
CC The sequences were used to generate the peptides AAR96424-R96524. The
CC sequences can also be used to synthesise probes and primers for the
CC detection of HCV in a sample. The polypeptides can be used to detect
CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.

SO Sequence 447 BP; 82 A; 130 C; 114 G; 118 T; 3 other;

Query Match 52.0%; Score 13; DB 17; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cagccagcagc 14
| | | | | | | | | |
DB 406 cagccagcagc 418

RESULT 43

AAC54077
ID AAC54077 standard; DNA: 591 BP.

AC AAC54077;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 76608.

KW Hybridisation assay; genetic mapping; gene expression control;
protein identification; signal transduction pathway;
metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 16-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0129845.

PR 21-APR-1999; 99US-0130077.

PR 23-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 04-MAY-1999; 99US-0132048.

PR 06-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 07-MAY-1999; 99US-0132486.

PR 11-MAY-1999; 99US-0132487.

PR 14-MAY-1999; 99US-0132488.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137724.

PR 07-JUN-1999; 99US-0138009.

PR 08-JUN-1999; 99US-0138504.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

```
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 22-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142877.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149729.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161933.
PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 591;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```


RESULT 44
AAH84601/C
ID AAH84601 standard; DNA: 597 BP.
XX
AC AAH84601:
XX
DT 26-SEP-2001 (first entry)
XX
DE E. coli growth and proliferation related coding sequence SEQ ID NO:229.
XX
KW Escherichia coli; growth; proliferation; microbial; antimicrobial;
XX bacterial infection; microorganism; ds.
XX
OS Escherichia coli.
XX
PN WO200134810-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30950.
XX
PR 09-NOV-1999; 99US-0164415.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Forsyth RA, Ohlsen K, Zyskind J;
XX
PI WPI: 2001-335933/35.
DR P-PSDB; AAG98930.
XX
PT Novel nucleic acids that inhibit Escherichia coli proliferation, useful
PT for screening for homologous genes and for designing expression vectors
PT
XX
PS Claim 9; Page 275-276; 522pp; English.
XX
XX AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
XX related DNA sequences (1). AAH84500 to AAH84670 encode the E. coli
XX growth and proliferation related proteins in AAG99078 and AAG98830
XX to AAG98999. (1) can be used as potential targets for the generation of
XX new antimicrobial agents, and for identification of compounds which
XX interact with the gene products of (1). In addition the expression of
XX (1) and the purification of the proteins, the purified proteins can be
XX used to generate reagents and screen small molecule libraries or other
XX candidate compound libraries for compounds that can be further developed
XX to yield novel antimicrobial compounds. In addition, nucleic acid probes
XX complementary to (1) that are specific for particular species of
XX microorganisms can be used to identify particular microorganism species
XX in clinical specimens, therefore, providing a rapid and dependable
XX method by which to identify the causative agents of a bacterial
XX infection. Also, antibodies generated against proteins translated from
XX mRNA transcribed from proliferation-required sequences can also be used
XX to screen for specific microorganisms that produce such proteins in a
XX species-specific manner. AAH84371 and AAH84670 represent sequencing
XX primers used in the isolation of E. coli growth and proliferation
XX related sequence, which are used in an example from the present
XX invention.
XX
SQ Sequence 597 BP; 130 A; 158 C; 184 G; 125 T; 0 other;

XX
AC AAC35502;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 10429.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
XX
PR 05-MAR-1999; 99US-0123180.
XX
PR 09-MAR-1999; 99US-0123548.
XX
PR 23-MAR-1999; 99US-0125788.
XX
PR 25-MAR-1999; 99US-0126264.
XX
PR 29-MAR-1999; 99US-0126785.
XX
PR 01-APR-1999; 99US-0127462.
XX
PR 06-APR-1999; 99US-0128234.
XX
PR 08-APR-1999; 99US-0128714.
XX
PR 16-APR-1999; 99US-0129845.
XX
PR 19-APR-1999; 99US-0130077.
XX
PR 21-APR-1999; 99US-0130449.
XX
PR 23-APR-1999; 99US-0130510.
XX
PR 28-APR-1999; 99US-0130891.
XX
PR 30-APR-1999; 99US-0132048.
XX
PR 30-APR-1999; 99US-0132407.
XX
PR 04-MAY-1999; 99US-0132484.
XX
PR 05-MAY-1999; 99US-0132485.
XX
PR 06-MAY-1999; 99US-0132486.
XX
PR 06-MAY-1999; 99US-0132487.
XX
PR 07-MAY-1999; 99US-0132863.
XX
PR 11-MAY-1999; 99US-0134256.
XX
PR 14-MAY-1999; 99US-0134218.
XX
PR 14-MAY-1999; 99US-0134219.
XX
PR 14-MAY-1999; 99US-0134221.
XX
PR 14-MAY-1999; 99US-0134370.
XX
PR 18-MAY-1999; 99US-0134768.
XX
PR 19-MAY-1999; 99US-0134941.
XX
PR 20-MAY-1999; 99US-0135124.
XX
PR 21-MAY-1999; 99US-0135353.
XX
PR 24-MAY-1999; 99US-0135629.
XX
PR 25-MAY-1999; 99US-0136021.
XX
PR 27-MAY-1999; 99US-0136332.
XX
PR 28-MAY-1999; 99US-0136782.
XX
PR 01-JUN-1999; 99US-0137222.
XX
PR 03-JUN-1999; 99US-0137528.
XX
PR 04-JUN-1999; 99US-0137502.
XX
PR 07-JUN-1999; 99US-0137724.
XX
PR 08-JUN-1999; 99US-0138094.
XX
PR 10-JUN-1999; 99US-0138540.
XX
PR 10-JUN-1999; 99US-0138847.
XX
PR 14-JUN-1999; 99US-0139119.
XX
PR 16-JUN-1999; 99US-0139452.
XX
PR 16-JUN-1999; 99US-0139453.
XX
PR 17-JUN-1999; 99US-0139492.
XX
PR 18-JUN-1999; 99US-0139454.
XX
PR 18-JUN-1999; 99US-0139455.
XX
PR 18-JUN-1999; 99US-0139456.
XX
PR 18-JUN-1999; 99US-0139457.
XX
PR 18-JUN-1999; 99US-0139458.
XX
PR 18-JUN-1999; 99US-0139459.
XX
PR 18-JUN-1999; 99US-0139460.
XX
PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999: 99US-0139462.
PR 18-JUN-1999: 99US-0139463.
PR 18-JUN-1999: 99US-0139750.
PR 18-JUN-1999: 99US-0139763.
PR 21-JUN-1999: 99US-0139817.
PR 22-JUN-1999: 99US-0139899.
PR 23-JUN-1999: 99US-0140353.
PR 23-JUN-1999: 99US-0140354.
PR 24-JUN-1999: 99US-0140695.
PR 28-JUN-1999: 99US-0140823.
PR 29-JUN-1999: 99US-0140991.
PR 30-JUN-1999: 99US-0141287.
PR 01-JUL-1999: 99US-0141842.
PR 01-JUL-1999: 99US-0142154.
PR 02-JUL-1999: 99US-0142055.
PR 06-JUL-1999: 99US-0142390.
PR 08-JUL-1999: 99US-0142803.
PR 09-JUL-1999: 99US-0142920.
PR 12-JUL-1999: 99US-0142977.
PR 13-JUL-1999: 99US-0143542.
PR 14-JUL-1999: 99US-0143624.
PR 15-JUL-1999: 99US-0144005.
PR 16-JUL-1999: 99US-0144085.
PR 16-JUL-1999: 99US-0144086.
PR 19-JUL-1999: 99US-0144325.
PR 19-JUL-1999: 99US-0144331.
PR 19-JUL-1999: 99US-0144332.
PR 19-JUL-1999: 99US-0144333.
PR 19-JUL-1999: 99US-0144334.
PR 19-JUL-1999: 99US-0144335.
PR 20-JUL-1999: 99US-0144352.
PR 20-JUL-1999: 99US-0144632.
PR 20-JUL-1999: 99US-0144884.
PR 21-JUL-1999: 99US-0144814.
PR 21-JUL-1999: 99US-0145086.
PR 21-JUL-1999: 99US-0145088.
PR 22-JUL-1999: 99US-0145085.
PR 22-JUL-1999: 99US-0145087.
PR 22-JUL-1999: 99US-0145089.
PR 22-JUL-1999: 99US-0145192.
PR 23-JUL-1999: 99US-0145145.
PR 23-JUL-1999: 99US-0145218.
PR 23-JUL-1999: 99US-0145224.
PR 26-JUL-1999: 99US-0145276.
PR 27-JUL-1999: 99US-0145913.
PR 27-JUL-1999: 99US-0145916.
PR 27-JUL-1999: 99US-0145919.
PR 28-JUL-1999: 99US-0145951.
PR 02-AUG-1999: 99US-0146386.
PR 02-AUG-1999: 99US-0146388.
PR 03-AUG-1999: 99US-0146389.
PR 03-AUG-1999: 99US-0147038.
PR 04-AUG-1999: 99US-0147204.
PR 04-AUG-1999: 99US-0147302.
PR 05-AUG-1999: 99US-0147192.
PR 05-AUG-1999: 99US-0147260.
PR 06-AUG-1999: 99US-0147303.
PR 06-AUG-1999: 99US-0147416.
PR 09-AUG-1999: 99US-0147493.
PR 09-AUG-1999: 99US-0147935.
PR 10-AUG-1999: 99US-0148171.
PR 11-AUG-1999: 99US-0148319.
PR 12-AUG-1999: 99US-0148341.
PR 13-AUG-1999: 99US-0148565.
PR 13-AUG-1999: 99US-0148684.
PR 16-AUG-1999: 99US-0149368.
PR 17-AUG-1999: 99US-0149375.
PR 18-AUG-1999: 99US-0149426.
PR 20-AUG-1999: 99US-0149722.
PR 20-AUG-1999: 99US-0149723.
PR 23-AUG-1999: 99US-0149902.
PR 23-AUG-1999: 99US-0149902.
PR 23-AUG-1999: 99US-0149930.

PR 25-AUG-1999: 99US-0150566.
PR 26-AUG-1999: 99US-0150884.
PR 27-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151066.
PR 27-AUG-1999: 99US-0151080.
PR 30-AUG-1999: 99US-0151303.
PR 31-AUG-1999: 99US-0151438.
PR 01-SEP-1999: 99US-0151930.
PR 07-SEP-1999: 99US-0152363.
PR 10-SEP-1999: 99US-0153070.
PR 13-SEP-1999: 99US-0153758.
PR 15-SEP-1999: 99US-0154018.
PR 16-SEP-1999: 99US-0154039.
PR 20-SEP-1999: 99US-0154779.
PR 22-SEP-1999: 99US-0155139.
PR 23-SEP-1999: 99US-0155486.
PR 24-SEP-1999: 99US-0155659.
PR 28-SEP-1999: 99US-0156458.
PR 29-SEP-1999: 99US-0156596.
PR 04-OCT-1999: 99US-0157117.
PR 05-OCT-1999: 99US-0157173.
PR 06-OCT-1999: 99US-0157865.
PR 07-OCT-1999: 99US-0158029.
PR 08-OCT-1999: 99US-0158232.
PR 12-OCT-1999: 99US-0158369.
PR 13-OCT-1999: 99US-0159293.
PR 13-OCT-1999: 99US-0159294.
PR 13-OCT-1999: 99US-0159295.
PR 14-OCT-1999: 99US-0159329.
PR 14-OCT-1999: 99US-0159330.
PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159637.
PR 14-OCT-1999: 99US-0159638.
PR 18-OCT-1999: 99US-0159584.
PR 21-OCT-1999: 99US-0160741.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 21-OCT-1999: 99US-0160770.
PR 21-OCT-1999: 99US-0160814.
PR 21-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160980.
PR 22-OCT-1999: 99US-0160981.
PR 22-OCT-1999: 99US-0160989.
PR 22-OCT-1999: 99US-0161404.
PR 25-OCT-1999: 99US-0161405.
PR 25-OCT-1999: 99US-0161406.
PR 25-OCT-1999: 99US-0161406.
PR 26-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161922.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.

Query Match 52.0%; Score 13; DB 21; length 606;
Best local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 12 attcgaccccg 24
Db 159 attcgaccccg 171

Search completed: December 26, 2001, 14:16:57
Job time: 7009 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:55:29 ; Search time 143.1 Seconds
(without alignments)
39.566 Million cell updates/sec

Title: US-09-396-196f-7

Perfect score: 25

Sequence: 1 gcatgcgcagcattcgcctcgcgt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Capext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 9

Total number of hits satisfying chosen parameters: 9225

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : Issued_Patents.NA.*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTCUTS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	14	56.0	2232	3	PCT-US96-05320A-113
6	14	56.0	4223	4	US-09-541-782-5
7	14	56.0	5319	1	US-08-169-927-1
8	13	52.0	61	3	US-09-023-228B-59
9	13	52.0	447	4	US-08-836-075A-51
10	13	52.0	1460	2	US-08-933-750C-60
11	13	52.0	1460	3	US-09-234-613-60
12	13	52.0	1650	2	US-08-743-637B-172
13	13	52.0	1650	3	US-08-526-840B-112
14	13	52.0	2104	1	US-08-682-193A-1
15	13	52.0	2193	1	US-08-420-235B-16
16	13	52.0	2193	3	US-08-793-624A-16
17	13	52.0	2193	5	PCT-US95-10194-16
18	13	52.0	2312	2	US-09-134-566-15
19	13	52.0	2314	2	US-09-134-566-8
20	13	52.0	2973	2	US-08-343-101A-5
21	13	52.0	2973	3	US-09-183-688-5
22	13	52.0	20710	1	US-08-420-235B-1
23	13	52.0	20710	3	US-08-793-624A-1
24	13	52.0	20710	5	PCT-US95-10194-1
25	13	52.0	35100	2	US-08-770-379-18
26	13	52.0	35100	4	US-08-757-669A-18
27	12	48.0	19	3	US-08-851-843A-92

28	12	48.0	19	4	US-08-974-549A-384	Sequence 384, App
29	12	48.0	19	4	US-08-854-050-92	Sequence 92, Appl
30	12	48.0	38	1	US-08-373-124A-1692	Sequence 1692, Ap
31	12	48.0	38	1	US-08-435-628-1692	Sequence 1692, Ap
32	12	48.0	660	2	US-08-743-637B-175	Sequence 175, App
33	12	48.0	660	3	US-08-526-840B-175	Sequence 175, App
34	12	48.0	715	4	US-08-998-416-767	Sequence 767, App
35	12	48.0	780	3	US-08-927-1-8-3	Sequence 3, Appl
36	12	48.0	841	4	US-08-960-823-39	Sequence 39, Appl
37	12	48.0	864	2	US-08-595-868B-11	Sequence 11, Appl
38	12	48.0	867	4	US-09-139-819A-11	Sequence 11, Appl
39	12	48.0	954	1	US-08-624-125-9	Sequence 9, Appl
40	12	48.0	1058	4	US-08-818-112-45	Sequence 45, Appl
41	12	48.0	1188	1	US-08-781-562-2	Sequence 2, Appl
42	12	48.0	1239	4	US-09-064-693A-20	Sequence 20, Appl
43	12	48.0	1249	3	US-08-985-950-1	Sequence 1, Appl
44	12	48.0	1305	2	US-08-484-575A-19	Sequence 19, Appl
45	12	48.0	1305	3	US-08-477-459-19	Sequence 19, Appl
46	12	48.0	1305	3	US-08-479-869-19	Sequence 19, Appl
47	12	48.0	1305	3	US-08-486-414-19	Sequence 19, Appl
48	12	48.0	1305	5	PCT-US94-01826A-19	Sequence 19, Appl
49	12	48.0	1305	5	PCT-US94-02252A-19	Sequence 19, Appl
50	12	48.0	1305	5	PCT-US96-03916-10	Sequence 10, Appl
51	12	48.0	1411	1	US-07-952-817-10	Sequence 10, Appl
52	12	48.0	1411	6	5210025-3	Patent No. 5210025
53	12	48.0	1514	1	US-08-207-900B-1	Sequence 1, Appl
54	12	48.0	1779	5	PCT-US96-05320A-831	Sequence 831, App
55	12	48.0	2171	3	US-08-851-843A-100	Sequence 100, App
56	12	48.0	2171	4	US-08-974-549A-266	Sequence 266, App
57	12	48.0	2171	4	US-08-854-050-100	Sequence 100, App
58	12	48.0	2176	4	US-08-974-549A-3	Sequence 3, Appl
59	12	48.0	2230	1	US-08-217-327-5	Sequence 5, Appl
60	12	48.0	2876	4	US-09-134-607A-9	Sequence 9, Appl
61	12	48.0	2876	4	US-09-134-607A-14	Sequence 14, Appl
62	12	48.0	2876	4	US-09-134-607A-11	Sequence 11, Appl
63	12	48.0	2938	4	US-08-969-815-3	Sequence 3, Appl
64	12	48.0	2938	4	US-09-120-025-3	Sequence 3, Appl
65	12	48.0	3107	1	US-07-846-181-4	Sequence 4, Appl
66	12	48.0	3107	1	US-07-845-989-4	Sequence 4, Appl
67	12	48.0	3265	4	US-09-134-607A-15	Sequence 15, Appl
68	12	48.0	3505	1	US-08-454-097-34	Sequence 34, Appl
69	12	48.0	3505	1	US-08-468-036-43	Sequence 43, Appl
70	12	48.0	3505	2	US-08-376-843-43	Sequence 43, Appl
71	12	48.0	3505	3	US-08-185-359-34	Sequence 34, Appl
72	12	48.0	3697	1	US-08-571-758-1	Sequence 1, Appl
73	12	48.0	3697	1	US-08-909-984A-1	Sequence 1, Appl
74	12	48.0	3697	1	US-08-909-983-1	Sequence 1, Appl
75	12	48.0	3855	4	US-08-974-549A-4	Sequence 4, Appl
76	12	48.0	3905	1	US-08-452-083-1	Sequence 1, Appl
77	12	48.0	4015	3	US-08-851-843A-224	Sequence 224, App
78	12	48.0	4015	4	US-08-974-549A-1	Sequence 1, Appl
79	12	48.0	4025	4	US-08-854-050-224	Sequence 224, App
80	12	48.0	4025	4	US-08-851-843A-173	Sequence 173, App
81	12	48.0	4029	4	US-08-974-549A-292	Sequence 292, App
82	12	48.0	4037	4	US-08-854-050-173	Sequence 173, App
83	12	48.0	4037	4	US-08-974-549A-343	Sequence 343, App
84	12	48.0	4147	6	5223610-4	Patent No. 5223610
85	12	48.0	4177	6	US-09-023-082A-23	Sequence 23, Appl
86	12	48.0	4208	6	5244657-4	Patent No. 5244657
87	12	48.0	4208	6	5433945-4	Patent No. 5433945
88	12	48.0	4530	4	US-09-064-693A-26	Sequence 26, Appl
89	12	48.0	8533	1	US-07-846-181-6	Sequence 6, Appl
90	12	48.0	8533	1	US-07-845-989-6	Sequence 6, Appl
91	12	48.0	11219	1	US-07-642-734C-1	Sequence 1, Appl
92	12	48.0	11219	3	US-08-439-009A-1	Sequence 1, Appl
93	12	48.0	13473	5	PCT-US96-03916-59	Sequence 59, Appl
94	12	48.0	18912	5	US-07-642-734C-3	Sequence 3, Appl
95	12	48.0	20235	5	US-08-439-009A-3	Sequence 3, Appl
96	12	48.0	20235	4	US-09-144-085-3	Sequence 3, Appl
97	12	48.0	33529	3	US-08-923-1137-C	Sequence 2, Appl
98	12	48.0	36519	2	US-08-804-227C-1	Sequence 1, Appl
99	12	48.0	43380	2	US-08-103-840A-2	Sequence 2, Appl
100	12	48.0	4403765	4		

ALIGNMENTS

```
RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5839335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product="Biotin synthase"
; OTHER INFORMATION: /evidence=EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 1.le-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgcacgacattgcattcctcgt 25
    |||
Db 90 GCATGCCACGACATTGCATCTCTCGT 114

RESULT 2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
```

```
ADDRESS: No. 5869719artle Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product="Biotin synthase"
; OTHER INFORMATION: /evidence=EXPERIMENTAL
US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 1.le-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgcacgacattgcattcctcgt 25
    |||
Db 90 GCATGCCACGACATTGCATCTCTCGT 114

RESULT 3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: PB030A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 117
OTHER INFORMATION: /product= "biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biob"
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "involved in pimeoyl-CoA synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "bioc"
OTHER INFORMATION: /number= 3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3750
OTHER INFORMATION: /ec_number= 2.6.1.62
OTHER INFORMATION: /product= "DAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioA"
OTHER INFORMATION: /number= 5
OTHER INFORMATION: /standard_name= "S-Adenosyl-L-methionine:8-amino-7-oxononanoate"
OTHER INFORMATION: /aminoacids= 1
FEATURE:
NAME/KEY: CDS
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 5098
OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ORF1"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: -10-signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "promoter plac"
FEATURE:
NAME/KEY: -35-signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name= "promoter plac"

FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "biob RBS no. 9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name= "bioA RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name= "ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name= "rho-independent"
OTHER INFORMATION: /transcriptional terminator
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "promoter plac"
OTHER INFORMATION: /evidence= EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. NO. 1.1e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatccacgattcgcctcgt 25
|||||
Db 206 GCATCCACGATTTGCATCCTCGT 230

RESULT 4
US-08-411-768B-6
Sequence 6, Application US/08411768B
Patent No. 6083712
GENERAL INFORMATION:
APPLICANT: Olwen Birch
APPLICANT: Johann Brass
APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
SOFTWARE: Version 5.1

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: PB030A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=1154
OTHER INFORMATION: /EC_number=2.3.1.47
OTHER INFORMATION: /product="KAPA synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="biof"
OTHER INFORMATION: /number=2
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=3043
OTHER INFORMATION: /EC_number=6.3.3.3
OTHER INFORMATION: /product="PTB synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="biof"
OTHER INFORMATION: /number=4
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name="biof RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name="biof RBS"
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6

Query Match      100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1,1e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  gcatgccagcattcgcattcgt 25
         |||||||
Db      206  GCATGCCAGCATTTCGATCTCGT 230

RESULT      5
PCT-US96-05320A-113/c
; Sequence 113, Application PC/TUS9605320A
; GENERAL INFORMATION:
```

```

APPLICANT: Human Genome Sciences
APPLICANT: 9410 Key West Avenue
APPLICANT: Rockville, MD 20850
APPLICANT: United States of America
APPLICANT: Johns Hopkins University
APPLICANT: 720 Rutland Avenue
APPLICANT: Baltimore, MD 21205
APPLICANT: United States of America
APPLICANT: Mark D. Adams
APPLICANT: Owen White
APPLICANT: Hamilton O. Smith
APPLICANT: J. Craig Venter
TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20003-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/05320A
FILING DATE: April 22, 1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: June 7, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: June 7, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Eric K. Steffe
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488, 014PC01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 2232 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
PCT-US96-05320A-113

Query Match      56.0%; Score 14; DB 5; Length 2232;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3  atgccagcattc 16
         |||||||
Db      636  ATGCCAGCATTTC 623

RESULT      6
US-09-541-782-5/c
; Sequence 5, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Berard, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/541,782
; CURRENT FILING DATE: 2000-04-03
```

NUMBER OF SEQ ID NOS: 10
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 5
LENGTH: 4223
TYPE: DNA
ORGANISM: S.pombe
US-09-541-782-5

Query Match 56.0%; Score 14; DB 4; Length 4223;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccacattt 15
|||||
Db 1559 CATGCCACATT 1546

RESULT 7
US-08-169-927-1/C
Sequence 1, Application US/08169927
Patent No. 5783441

GENERAL INFORMATION:

APPLICANT: Carl, Mitchell
APPLICANT: Dobson, Michael E.
APPLICANT: Ching, Wei Wei
APPLICANT: Dasch, Gregory A.
TITLE OF INVENTION: Gene and Protein Applicable to the
TITLE OF INVENTION: Preparation of Vaccines for Rickettsia typhi and the Detection of Both
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Counsel, Naval Medical R & D Command
STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
CITY: Bethesda
STATE: MD
COUNTRY: USA
ZIP: 20889-5606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/169,927
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/742,128
FILING DATE: 08/09/91

ATTORNEY/AGENT INFORMATION:
NAME: Spevack, A. David

REGISTRATION NUMBER: 24,743
REFERENCE/DOCKET NUMBER: 75,976
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-6759

TELEFAX: (301) 295-1022
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 5319 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal

ORIGINAL SOURCE:

ORGANISM: Rickettsia prowazekii

STRAIN: Breinl

FEATURE:

NAME/KEY: -35_signal

LOCATION: 340..345

FEATURE:

NAME/KEY: -10_signal

LOCATION: 363..368

FEATURE:

NAME/KEY: CDS

LOCATION: 391..5226

FEATURE:

NAME/KEY: RBS

LOCATION: 379..386

FEATURE:

NAME/KEY: stem_loop

LOCATION: 5270..5306

PUBLICATION INFORMATION:

AUTHORS: Carl, M.

AUTHORS: Dobson, M. E.

AUTHORS: Ching, W. M.

AUTHORS: Dasch, G. A.

TITLE: Characterization of the gene encoding the

TITLE: protective S-layer protein of Rickettsia

TITLE: prowazekii; presence of a truncated identical

TITLE: homolog in rickettsia typhi

JOURNAL: Proc. Natl. Acad. Sci. U.S.A.

DATE: 1990

RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 5319

US-08-169-927-1

Query Match

Best Local Similarity 56.0%; Score 14; DB 1; Length 5319;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccacattt 15
|||||
Db 2221 CATGCCACATT 2208

RESULT 8
US-09-023-228B-59

Sequence 59, Application US/09023228B
Patent No. 6140490

GENERAL INFORMATION:

APPLICANT: BIESSECKER, GREGORY

APPLICANT: GOLD, LARRY

TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS OF

TITLE OF INVENTION: COMPLEMENT SYSTEM PROTEINS

NUMBER OF SEQUENCES: 157

CORRESPONDENCE ADDRESS:

ADDRESSEE: Swanson & Bratschun, L.L.C.

STREET: 8400 E. Prentice Place #200

CITY: Denver

STATE: Colorado

COUNTRY: US

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WordPerfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/023,228B

FILING DATE: 12-FEBRUARY-1998

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US97/011739

FILING DATE: 30 JAN 1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/595,335

FILING DATE: 1 FEB 1996

ATTORNEY/AGENT INFORMATION:

NAME: Barry J. Swanson, Esq.

REGISTRATION NUMBER: 33,215

REFERENCE/DOCKET NUMBER: NEX50/CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 793-3333

TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: RNA
; FEATURE:
; OTHER INFORMATION: All pyrimidines are 2'-F modified
US-09-023-228b-59

Query Match 52.0%; Score 13; DB 3; Length 61;
Best Local Similarity 69.2%; Pred. No. 55;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 8 ccagcattcgatc 20
Db 23 CAGCAUUCGAGC 35

RESULT 9
US-08-836-075A-51
; Sequence 51, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,075A
; FILING DATE: 21 Apr 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04155
; FILING DATE: 23 Oct 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-836-075A-51

Query Match 52.0%; Score 13; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 catgcgcagcatt 14
Db 406 CATGCAGCAGATT 418

RESULT 10
US-08-933-750C-60
; Sequence 60, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preethi
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSP for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750C
; FILING DATE: September 23, 1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1460 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: T14MOR01
; CLONE: 140516
US-08-933-750C-60

Query Match 52.0%; Score 13; DB 2; Length 1460;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 ccagcattcgat 19
Db 722 CCAGCATTCGAT 734

RESULT 11

US-09-234-613-60
; Sequence 60, Application US/09234613
; Patent No. 6132973
; GENERAL INFORMATION:
; APPLICANT: Lal, Pirelli
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/234,613
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750
; FILING DATE: September 23, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1460 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TLYMGR01
; CLONE: 140516
; US-09-234-613-60

Query Match 52.0%; Score 13; DB 3; Length 1460;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ccagcattcgatc 19
|||||
Db 722 CCAGCATTTCGAT 734

RESULT 12
US-08-743-637B-172/C
; Sequence 172, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND

; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 EAST WISCONSIN AVENUE
; CITY: MILWAUKEE
; STATE: WISCONSIN
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/743,637B
; FILING DATE: 04-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,840
; FILING DATE: 11-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586,90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5500
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 172:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1650 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-743-637B-172

Query Match 52.0%; Score 13; DB 2; Length 1650;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 cagcattcgatc 20
|||||
Db 555 CAGCATTTCGATC 543

RESULT 13
US-08-526-840B-172/C
; Sequence 172, Application US/08526840B
; Patent No. 6001564
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
; TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/526,840B
: FILING DATE: 11-SEP-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/304,732
: FILING DATE: 12-SEP-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: BAKER, Jean C.
: REGISTRATION NUMBER: 35,433
: REFERENCE/DOCKET NUMBER: 850586,90012
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (414) 277-5000
: TELEFAX: (414) 277-5591
: INFORMATION FOR SEQ ID NO: 172:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1650 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-526-840B-172

```

```

Query Match          52.0%; Score 13; DB 3; Length 1650;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      8 cagcattcgcac 20
      |||
Db      555 CAGCATTTCGATC 543

```

```

RESULT 14
US-08-682-193A-1
: Sequence 1, Application US/08682193A
: Patent No. 5776740
: GENERAL INFORMATION:
: APPLICANT: HATAKEYAMA, Kazuhisa
: APPLICANT: GOTO, Makoto
: APPLICANT: TERASAWA, Masato
: APPLICANT: YUKAWA, Hideaki
: TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
: STREET: 130 WATER STREET
: CITY: BOSTON
: STATE: MA
: COUNTRY: USA
: ZIP: 02019
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ Version 1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/682,193A
: FILING DATE: 17-JUL-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 7-181730
: FILING DATE: 18-JUL-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: DAVID, RESNICK S
: REGISTRATION NUMBER: 34,235
: REFERENCE/DOCKET NUMBER: 46643
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-523-6440
: TELEFAX: 200291 STRE
: INFORMATION FOR SEQ ID NO: 1:

```

```

: SEQUENCE CHARACTERISTICS:
: LENGTH: 2104 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE:
: ORIGINAL SOURCE:
: ORGANISM: Brevibacterium flavum
: STRAIN: MJ-233
: FEATURE:
: NAME/KEY: Coding Sequence
: LOCATION: 556..1855
: OTHER INFORMATION:
US-08-682-193A-1

```

```

Query Match          52.0%; Score 13; DB 1; Length 2104;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      13 ttctgactcgtc 25
      |||
Db      1635 TTTCGATCCTCGT 1647

```

```

RESULT 15
US-08-420-235B-16/c
: Sequence 16, Application US/08420235B
: Patent No. 5801042
: GENERAL INFORMATION:
: APPLICANT: Chang, Yuan
: APPLICANT: Moore, Patrick S.
: TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
: NUMBER OF SEQUENCES: 47
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cooper & Dunham LLP
: STREET: 1185 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/420,235B
: FILING DATE:
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 45185-B
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 278-0400
: TELEFAX: (212) 391-0525
: INFORMATION FOR SEQ ID NO: 16:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2193 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: N
: ANTI-SENSE: N
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..2193

```

OTHER INFORMATION:
US-08-420-235B-16

Query Match
Best Local Similarity 52.0%; Score 13; DB 1; Length 2193;
Pred. No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcatcgccagcat 13
|||||
DB 45 GCATCGCCAGCAT 33

RESULT 16
US-08-793-624-16/C
Sequence 16, Application US/08793624C
Patent No. 6150093
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And
FILE REFERENCE: 45185-C-PCT-US/JPM
CURRENT APPLICATION NUMBER: US/08/793,624C
CURRENT FILING DATE: 1997-02-18
NUMBER OF SEQ ID NOS: 58
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 16
LENGTH: 2193
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-08-793-624-16

Query Match
Best Local Similarity 52.0%; Score 13; DB 3; Length 2193;
Pred. No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcatcgccagcat 13
|||||
DB 45 GCATCGCCAGCAT 33

RESULT 17
PCT-US95-10194-16/C
Sequence 16, Application PC/TUS9510194
GENERAL INFORMATION:
APPLICANT: The Trustees of Columbia University in the City of New York
APPLICANT: City
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10194
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPM/MS
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SKD ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2193 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2193
OTHER INFORMATION:
PCT-US95-10194-16

Query Match
Best Local Similarity 52.0%; Score 13; DB 5; Length 2193;
Pred. No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcatcgccagcat 13
|||||
DB 45 GCATCGCCAGCAT 33

RESULT 18
US-09-134-566-15
Sequence 15, Application US/09134566
Patent No. 5998147
GENERAL INFORMATION:
APPLICANT: Denoyelle, Jysson, Francoise
APPLICANT: Petit, Christine
APPLICANT: Weill, Dominique
APPLICANT: Marlin-Duvernois, Sandrine
TITLE OF INVENTION: MUTATED POLYNUCLEOTIDE CORRESPONDING TO A MUTATION
TITLE OF INVENTION: RESPONSIBLE FOR PRELINGUAL, NON-SYNDROMIC DEAFNESS IN
TITLE OF INVENTION: THE CONNEXIN 26 GENE AND METHOD OF DETECTING THIS
TITLE OF INVENTION: HEREDITARY DEFECT
FILE REFERENCE: 03495-0175
CURRENT APPLICATION NUMBER: US/09/134,566
CURRENT FILING DATE: 1998-08-14
EARLIER APPLICATION NUMBER: 60/055,863
EARLIER FILING DATE: 1998-08-15
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 15
LENGTH: 2312
TYPE: DNA
ORGANISM: human connexin
US-09-134-566-15

Query Match
Best Local Similarity 52.0%; Score 13; DB 2; Length 2312;
Pred. No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tcgcagcatlcc 16
|||||
DB 960 tcgcagcatlcc 972

RESULT 19
US-09-134-566-8
Sequence 8, Application US/09134566
Patent No. 5998147
GENERAL INFORMATION:
APPLICANT: Petit, Christine
APPLICANT: Denoyelle-Gryson, Francoise
APPLICANT: Weill, Dominique
APPLICANT: Marlin-Duvernois, Sandrine

APPLICANT: Guesdon, Jean-Luc
TITLE OF INVENTION: MUTATED POLYNUCLEOTIDE CORRESPONDING TO A MUTATION
TITLE OF INVENTION: RESPONSIBLE FOR PRELIMINARY NON-SYNDROMIC DEAFNESS IN
TITLE OF INVENTION: THE CONNEXIN 26 GENE AND METHOD OF DETECTING THIS
FILE REFERENCE: 03495-0175
CURRENT APPLICATION NUMBER: US/09/134,566
EARLIER FILING DATE: 1998-08-14
EARLIER APPLICATION NUMBER: 60/055,863
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 8
LENGTH: 2314
TYPE: DNA
ORGANISM: human connexin
US-09-134-566-8

Query Match 52.0%; Score 13; DB 2; Length 2314;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tcgcgcagcattc 16
|||||
DB 962 tcgcgcagcattc 974

RESULT 20
US-08-343-101A-5/c
Sequence 5, Application US/08343101A
Patent No. 5830759
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma
TITLE OF INVENTION: Virus Sequences And Uses Thereof
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,101A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: White Esq., John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0526
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2973 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
US-08-343-101A-5

Query Match 52.0%; Score 13; DB 2; Length 2973;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcatcgccagcat 13
|||||
DB 2181 GCATCGCCAGCAT 2169

RESULT 21
US-09-183-688-5/c
Sequence 5, Application US/09183688
Patent No. 6093550
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma
TITLE OF INVENTION: Virus Sequences And Uses Thereof
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/183,688
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/343,101
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: White Esq., John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0526
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2973 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
US-09-183-688-5

Query Match 52.0%; Score 13; DB 3; Length 2973;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcatcgccagcat 13
|||||
DB 2181 GCATCGCCAGCAT 2169

RESULT 22
US-08-420-235B-1/c
Sequence 1, Application US/08420235B
Patent No. 5801042
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma
TITLE OF INVENTION: Virus Sequences And Uses Thereof
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/420,235B
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/343,101
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: White Esq., John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0526
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2973 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
US-08-420-235B-1/c

TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
TITLE OF INVENTION: SEQUENCES AND USES THEREOF

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESSES:

ADDRESS: Cooper & Dunham LLP

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/420,235B

FILING DATE:

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 45185-B

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0525

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 20710 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: N

ANTI-SENSE: N

US-08-420-235B-1

Query Match

Best Local Similarity 52.0%; Score 13; DB 1; Length 20710;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatcgccagcat 13

Db 2137 GCATCGCCAGCAT 2125

RESULT 23

US-08-793-624-1/c

Sequence 1, Application US/08793624C

Patent No. 6150093

GENERAL INFORMATION:

APPLICANT: Chang, Yuan

APPLICANT: Moore, Patrick S.

TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And

TITLE OF INVENTION: Uses Thereof

FILE REFERENCE: 45185-C-PCT-US/JPW

CURRENT APPLICATION NUMBER: US/08/793,624C

CURRENT FILING DATE: 1997-02-18

NUMBER OF SEQ ID NOS: 58

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 1

LENGTH: 20710

TYPE: DNA

ORGANISM: Kaposi's sarcoma-associated herpesvirus

US-08-793-624-1

Query Match

Best Local Similarity 52.0%; Score 13; DB 3; Length 20710;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatcgccagcat 13

Db 2137 GCATCGCCAGCAT 2125

RESULT 23

US-08-793-624-1/c

Sequence 1, Application US/08793624C

Patent No. 6150093

GENERAL INFORMATION:

APPLICANT: Chang, Yuan

APPLICANT: Moore, Patrick S.

TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And

TITLE OF INVENTION: Uses Thereof

FILE REFERENCE: 45185-C-PCT-US/JPW

CURRENT APPLICATION NUMBER: US/08/793,624C

CURRENT FILING DATE: 1997-02-18

NUMBER OF SEQ ID NOS: 58

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 1

LENGTH: 20710

TYPE: DNA

ORGANISM: Kaposi's sarcoma-associated herpesvirus

US-08-793-624-1

Db 2137 GCATCGCCAGCAT 2125

RESULT 24

PCT-US95-10194-1/c

Sequence 1, Application PC/US9510194

GENERAL INFORMATION:

APPLICANT: The Trustees of Columbia University in the City of New York

APPLICANT: City

TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS

TITLE OF INVENTION: SEQUENCES AND USES THEREOF

NUMBER OF SEQUENCES: 45

CORRESPONDENCE ADDRESS:

ADDRESS: Cooper & Dunham LLP

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/10194

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPW/MSC

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0525

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 20710 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: N

ANTI-SENSE: N

PCT-US95-10194-1

Query Match

Best Local Similarity 52.0%; Score 13; DB 5; Length 20710;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatcgccagcat 13

Db 2137 GCATCGCCAGCAT 2125

RESULT 25

US-08-770-18/c

Sequence 18, Application US/08770379

Patent No. 5849564

GENERAL INFORMATION:

APPLICANT: Chang, Yuan

APPLICANT: Bohenzky, Roy A.

APPLICANT: Russo, James J.

APPLICANT: Edelman, Isidore S.

APPLICANT: Moore, Patrick S.

TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED

TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESS: Cooper & Dunham LLP

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/10194

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPW/MSC

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0525

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 20710 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: N

ANTI-SENSE: N

PCT-US95-10194-1

CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,379
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 35100 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-770-379-18

Query Match 52.0%; Score 13; DB 2; Length 35100;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcat 13
|||||

Db 2057 GCATGCCAGCAT 2045

RESULT 26
US-08-757-669A-18/c
Sequence 18, Application US/08757669A
Patent No. 6183751
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/757,669A
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-F
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 35100 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-757-669A-18

Query Match 52.0%; Score 13; DB 4; Length 35100;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcat 13
|||||

Db 2057 GCATGCCAGCAT 2045

RESULT 27
US-08-851-843A-92/c
Sequence 92, Application US/08851843A
Patent No. 6093809
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Hailey, Calvin
APPLICANT: Andrews, William H.
TITLE OF INVENTION: No. 6093809el Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,843A
FILING DATE: 06-MAY-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 92:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-851-843A-92

Query Match 48.0%; Score 12; DB 3; Length 19;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagcat 13
|||||
DB 19 CATGCCAGCAT 8

RESULT 28
US-08-974-549A-384/C
Sequence 384, Application US/08974549A
Patent No. 6166178
GENERAL INFORMATION:
APPLICANT: Cecch, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin B.
APPLICANT: Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 727
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-NOV-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997

PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17685
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002610US
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 384:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: -
LOCATION: 1..19
OTHER INFORMATION: /note="TCP1.6 primer"
US-08-974-549A-384

Query Match 48.0%; Score 12; DB 4; Length 19;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagcat 13
|||||
DB 19 CATGCCAGCAT 8

RESULT 29
US-08-854-050-92/C
Sequence 92, Application US/08854050
Patent No. 6261836
GENERAL INFORMATION:
APPLICANT: Cecch, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin
APPLICANT: Andrews, William H.
TITLE OF INVENTION: No. 6261836el Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/854,050
FILING DATE: 09-MAY-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419

FILING DATE: 18-APR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-00293005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 92:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-854-050-92

Query Match 48.0%; Score 12; DB 4; Length 19;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 catgccacat 13
|||||
DB 19 CATGCCACAT 8

RESULT 30
US-08-373-124A-1692/C
Sequence 1692, Application US/08373124A
Patent No. 5646042
GENERAL INFORMATION:
APPLICANT: Slinchcomb, Dan T.
APPLICANT: Draper, Kenneth
APPLICANT: McSwigen, James
APPLICANT: Jarvis, Thale
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
TREATMENT OF RESTENOSIS AND
TITLE OF INVENTION: CANCER USING RIBOZYMES
NUMBER OF SEQUENCES: 2627
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/373,124A
FILING DATE: January 13, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/245,466
FILING DATE: May 18, 1994
APPLICATION NUMBER: 08/192,943
FILING DATE: February 7, 1994
APPLICATION NUMBER: 07/987,132
FILING DATE: December 7, 1992
APPLICATION NUMBER: 07/936,422
FILING DATE: August 26, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Wardburg, Richard

REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 209/035
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1692:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-373-124A-1692

Query Match 48.0%; Score 12; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 gccagcatctg 17
|||||
DB 38 GCCAGCATCTG 27

RESULT 31
US-08-435-628-1692/C
Sequence 1692, Application US/08435628
Patent No. 5817796
GENERAL INFORMATION:
APPLICANT: Slinchcomb, Dan T.
APPLICANT: Draper, Kenneth
APPLICANT: McSwigen, James
APPLICANT: Jarvis, Thale
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
TREATMENT OF RESTENOSIS AND
TITLE OF INVENTION: CANCER USING RIBOZYMES
NUMBER OF SEQUENCES: 2627
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/435,628
FILING DATE: 05-MAY-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/373,124
FILING DATE: January 13, 1995
APPLICATION NUMBER: 08/245,466
FILING DATE: May 18, 1994
APPLICATION NUMBER: 08/192,943
FILING DATE: February 7, 1994
APPLICATION NUMBER: 07/987,132
FILING DATE: December 7, 1992
APPLICATION NUMBER: 07/936,422
FILING DATE: August 26, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Wardburg, Richard
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 209/035
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440

TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1692:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-435-628-1692

Query Match 48.0%; Score 12; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 gccagattcgat 17
|||||
DB 38 gccagatttcg 27

RESULT 32
US-08-743-637B-175/c

Sequence 175, Application US/08743637B
Patent No. 5994066

GENERAL INFORMATION:

APPLICANT: BERGERON, Michel G.
APPLICANT: PICARD, Francois J.

APPLICANT: OUELLETTE, Marc

APPLICANT: ROY, Paul H.

TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA

TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND

TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED

TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...

NUMBER OF SEQUENCES: 273

CORRESPONDENCE ADDRESSES:

ADDRESSEE: QUARLES & BRADY

STREET: 411 EAST WISCONSIN AVENUE

CITY: MILWAUKEE

STATE: WISCONSIN

COUNTRY: USA

ZIP: 53202-4497

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/743,637B

FILING DATE: 04-NOV-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/526,840

FILING DATE: 11-SEP-1995

ATTORNEY/AGENT INFORMATION:

NAME: BAKER, Jean C.

REGISTRATION NUMBER: 35,433

REFERENCE/DOCKET NUMBER: 850586,90012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (414) 277-5000

TELEFAX: (414) 277-5591

INFORMATION FOR SEQ ID NO: 175:

SEQUENCE CHARACTERISTICS:

LENGTH: 660 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-743-637B-175

Query Match 48.0%; Score 12; DB 2; Length 660;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 gccagattcgat 19
|||||
DB 627 gccagatttcgat 616

RESULT 33

US-08-526-840B-175/c

Sequence 175, Application US/08526840B

Patent No. 6001564

GENERAL INFORMATION:

APPLICANT: BERGERON, Michel G.

APPLICANT: OUELLETTE, Marc

APPLICANT: ROY, Paul H.

TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND

TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY

TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE

TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...

NUMBER OF SEQUENCES: 177

CORRESPONDENCE ADDRESSES:

ADDRESSEE: QUARLES & BRADY

STREET: 411 East Wisconsin Avenue

CITY: Milwaukee

STATE: Wisconsin

COUNTRY: USA

ZIP: 53202-4497

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/526,840B

FILING DATE: 11-SEP-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/304,732

FILING DATE: 12-SEP-1994

ATTORNEY/AGENT INFORMATION:

NAME: BAKER, Jean C.

REGISTRATION NUMBER: 35,433

REFERENCE/DOCKET NUMBER: 850586,90012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (414) 277-5000

TELEFAX: (414) 277-5591

INFORMATION FOR SEQ ID NO: 175:

SEQUENCE CHARACTERISTICS:

LENGTH: 660 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-526-840B-175

Query Match 48.0%; Score 12; DB 3; Length 660;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 gccagattcgat 19
|||||
DB 627 gccagatttcgat 616

RESULT 34

US-08-998-416-767

Sequence 767, Application US/08998416

Patent No. 6239264

GENERAL INFORMATION:

APPLICANT: Philippesen, Peter

APPLICANT: Pohlmann, Rainer

APPLICANT: Steiner, Sabine

APPLICANT: Mohr, Christine

APPLICANT: Wendland, Jurgen

APPLICANT: Knechtle, Philipp
APPLICANT: Redischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYYP11
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264rlls Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CCG1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 767:
SEQUENCE CHARACTERISTICS:
LENGTH: 715 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1495RP
US-08-998-416-767

Query Match 48.0%; Score 12; DB 4; Length 715;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 tgcgcagcattt 15
|||||
Db 356 TCGCAGCATT 367

RESULT 35
US-08-927-128-3
Sequence 3, Application US/08927128
Patent No. 6127150
GENERAL INFORMATION:
APPLICANT: Coolidge, Thomas
APPLICANT: Wagner, Fred
APPLICANT: ven Heeke, Gino
APPLICANT: Schuster, Sheldon
APPLICANT: Stout, Jay
APPLICANT: Wylie, Dwane
TITLE OF INVENTION: PURIFICATION DIRECTED CLOSING OF PEPTIDES
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 3100 No. 6127150west Center, 90 S. 7th Street
CITY: Minneapolis
STATE: MN
COUNTRY: U.S.A.
ZIP: 55402
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/927,128
FILING DATE: 05-SEP-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/680,004
FILING DATE: 15-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: Carter, Charles G
REGISTRATION NUMBER: 35,093
REFERENCE/DOCKET NUMBER: 8648.2USD1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612/332-5300
TELEFAX: 612/332-9081
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 780 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1..780
OTHER INFORMATION:
US-08-927-128-3

Query Match 48.0%; Score 12; DB 3; Length 780;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 14 ttcgattctcgt 25
|||||
Db 532 TTCGATCTCGT 543

RESULT 36
US-08-990-823-39
Sequence 39, Application US/08990823D
Patent No. 6228371
GENERAL INFORMATION:
APPLICANT: Nano, Francis
TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
TITLE OF INVENTION: Immunostimulatory Peptides
FILE REFERENCE: 49086
CURRENT APPLICATION NUMBER: US/08/990,823D
EARLIER FILING DATE: 1997-12-15
EARLIER APPLICATION NUMBER: US 96/10375
EARLIER FILING DATE: 1996-06-14
EARLIER APPLICATION NUMBER: 60/000,254
EARLIER FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 113
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 39
LENGTH: 841
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
NAME/KEY: Modified base
OTHER INFORMATION: n represents a or g or c or t/u
US-08-990-823-39

Query Match 48.0%; Score 12; DB 4; Length 841;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 gcatcgccagca 12
|||||
Db 370 gcatcgccagca 381

RESULT 37
US-08-595-868C-11
; Sequence 11, Application US/08595868C
; Patent No. 5962270
; GENERAL INFORMATION:
; APPLICANT: Wagner, Fred
; APPLICANT: Stout, Jay
; APPLICANT: Henriksen, Dennis
; APPLICANT: Partridge, Bruce
; APPLICANT: Holmquist, Bart
; APPLICANT: Frank, Julie
; TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN FRAGMENTS AND US
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 5962270 West Center, 90 S. 7th Street
; CITY: Minneapolis
; STATE: MN
; COUNTRY: U.S.A.
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/595,868C
; FILING DATE: 06-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Carter, Charles G
; REGISTRATION NUMBER: 35,093
; REFERENCE/DOCKET NUMBER: 8648.590S01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612/332-5300
; TELEFAX: 612/332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 864 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1..864
; OTHER INFORMATION:
; US-08-595-868C-11

Query Match 48.0%; Score 12; DB 2; Length 864;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 14 ttgcgtctcgt 25
|||||

Db 532 TTGCATCTCTGT 543

RESULT 38
US-09-139-819A-11
; Sequence 11, Application US/09139819A
; Patent No. 6251635
; GENERAL INFORMATION:
; APPLICANT: WAGNER, Fred W.
; APPLICANT: STOUT, Jay S.
; APPLICANT: HENRIKSEN, Dennis B.
; APPLICANT: PARTRIDGE, Bruce E.
; APPLICANT: HOLMQUIST, Bart
; APPLICANT: FRANK, Julie A.
; TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN
; TITLE OF INVENTION: FRAGMENTS AND USE THEREOF IN THE PREPARATION OF CALCITONIN
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/139,819A
; FILING DATE: 25-AUG-1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/595,868
; FILING DATE: 06-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 089187/0144
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 867 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..864
; US-09-139-819A-11

Query Match 48.0%; Score 12; DB 4; Length 867;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 ttgcgtctcgt 25
|||||
Db 532 TTGCATCTCTGT 543

RESULT 39
US-08-624-125-9/C
; Sequence 9, Application US/08624125
; Patent No. 574341
; GENERAL INFORMATION:
; APPLICANT: CUNNINGHAM JR., FRANCIS X.
; APPLICANT: SUN, ZAIREN

;; TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND
;; TITLE OF INVENTION: METABOLISM AND A SYSTEM FOR SCREENING SUCH GENES
;; NUMBER OF SEQUENCES: 21
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;; ADDRESSEE: P.C.
;; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
;; CITY: ARLINGTON
;; STATE: VA
;; COUNTRY: USA
;; ZIP: 22202
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/624,125
;; FILING DATE: 29-MAR-1996
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: KELBER, STEVEN B.
;; REGISTRATION NUMBER: 30,073
;; REFERENCE/DOCKET NUMBER: 2747-063-27
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 703-413-3000
;; TELEFAX: 703-413-2220
;; INFORMATION FOR SEQ. ID NO: 9:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 954 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; US-08-624-125-9

QY 2 catgccacgat 13
|||||
Db 663 CATGCCACGAT 652

RESULT 40
US-08-818-112-45/c
; Sequence 45, Application US/08818112
; Patent No. 6280969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twadzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/818,112
;; FILING DATE: 13-MAR-1997
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: MAKI, David J.
;; REGISTRATION NUMBER: 31,392
;; REFERENCE/DOCKET NUMBER: 210121.411c6
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (206) 622-4900
;; TELEFAX: (206) 682-6031
;; INFORMATION FOR SEQ. ID NO: 45:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1058 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-818-112-45

Query Match 48.0%; Score 12; DB 4; Length 1058;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 ttgatcctcgt 25
|||||
Db 16 TTGATCCTCGT 5

RESULT 41
US-08-781-562-2/c
; Sequence 2, Application US/08781562
; Patent No. 5763589
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,562
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0181 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ. ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1188 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Consensus

; CLONE: Consensus
US-08-781-562-2

Query Match 48.0%; Score 12; DB 1; Length 1188;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 cgcacgatttc 16
|||||
DB 72 CGCCAGCATTTTC 61

RESULT 42

US-09-064-693A-20/c
; Sequence 20, Application US/09064693A
; Patent No. 6210937
; GENERAL INFORMATION:
; APPLICANT: Ward, Thomas E.
; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY
; TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. Gary Goodson
; ADDRESS: INEL--Lockheed Martin Idaho
; STREET: P.O. Box 1625
; CITY: Idaho Falls
; STATE: Idaho
; COUNTRY: USA
; ZIP: 83415-3810
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: Toshiba Satellite Pro 72150CDS
; OPERATING SYSTEM: Windows95
; SOFTWARE: Word Perfect 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/064,693A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: W. Gary Goodson
; REGISTRATION NUMBER: 22,387
; REFERENCE/DOCKET NUMBER: LIT-PI-296
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (208)526-9469
; TELEFAX: (208)526-8339
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1239 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-064-693A-20

Query Match 48.0%; Score 12; DB 4; Length 1239;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 atcgacgattt 14
|||||
DB 1137 ATCGCAGCATTT 1126

RESULT 43

US-08-985-950-1
; Sequence 1, Application US/08985950
; Patent No. 6140076

; GENERAL INFORMATION:
; APPLICANT: Adema, Gosse Jan
; TITLE OF INVENTION: Isolated Mammalian Monocyte Cell Genes;
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,950
; FILING DATE: 05-DEC-1997
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,279
; FILING DATE: 21-MARCH-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,181
; FILING DATE: 16-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/032,252
; FILING DATE: 06-DEC-1996

; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0670K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)852-9196
; TELEFAX: (650)496-1204

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1249 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 154..1062
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 211..1062
; US-08-985-950-1

Query Match 48.0%; Score 12; DB 3; Length 1249;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 cgcacgatttc 16
|||||
DB 197 CGCCAGCATTTTC 208

RESULT 44

US-08-484-575A-19/c
; Sequence 19, Application US/08484575A
; Patent No. 5923558
; GENERAL INFORMATION:
; APPLICANT: Mark D. Cochran and David E. Junker
; TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses Thereof
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York

```

STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,575A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White Esq, John P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)278-0450
TELEFAX: (212)391-0525
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1305 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1305
US-08-484-575A-19

```

```

Query Match      48.0%; Score 12; DB 2; Length 1305;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ccagcatttgcga 18
   |||||||
Db 780 CCAGCATTTCGA 769

```

```

RESULT 45
US-08-477-459-19/c
; Sequence 19, Application US/08477459
; Patent No. 6001369
; GENERAL INFORMATION:
; APPLICANT: Mark D. Cochran
; TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,459
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White Esq, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:

```

```

LENGTH: 1305 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1305
US-08-477-459-19

```

```

Query Match      48.0%; Score 12; DB 3; Length 1305;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ccagcatttgcga 18
   |||||||
Db 780 CCAGCATTTCGA 769

```

Search completed: December 26, 2001, 12:55:33
 Job time: 6965 sec

86 14 56.0 492 10 AW331662
 87 14 56.0 501 10 A1941630
 88 14 56.0 502 10 BE363404
 89 14 56.0 505 11 C98143
 90 14 56.0 513 10 AW746384
 91 14 56.0 514 11 BG467085
 92 14 56.0 514 11 B1166828
 93 14 56.0 515 10 AW174211
 94 14 56.0 515 13 A2261949
 95 14 56.0 516 10 A1834660
 96 14 56.0 517 10 A1534282
 97 14 56.0 521 11 BG102531
 98 14 56.0 522 13 A2427454
 99 14 56.0 524 10 A1957283
 100 14 56.0 526 13 A0882773

ALIGNMENTS

RESULT 1
 A1532390/c 498 bp mRNA EST 23-APR-2001
 LOCUS SD03891.5prime SD Drosophila melanogaster Schneider L2 cell culture
 DEFINITION PORT2 Drosophila melanogaster cDNA clone SD03891.5 similar to
 CG8473: Fban0008473 'transporter' located on: X 13E18-13P11;
 04/13/2001, mRNA sequence.
 A1532390
 VERSION A1532390.1 GI:4446525
 KEYWORDS EST.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 498)
 Harvey, D., Broksstein, P., Hong, L., Evans-Hoim, M., Su, C., Tsang, G.,
 Lewis, S. and Rubin, G.M.
 BDGP/HIMI Drosophila EST Project
 Unpublished (2001)
 Other ESTs: SD03891.3prime
 Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
 hit genomic AEC03500: arm: X 115446919,157763641
 estimated-cyto:13E14-14A4: 04/13/2001
 Plate: SD.38 row: H column: 7
 High quality sequence stop: 457
 POLYA-No.

FEATURES

source 1..498 Location/Qualifiers
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="SD03891"
 /clone_lib="SD Drosophila melanogaster Schneider L2 cell
 culture PORT2"
 /lab_host="DH5-alpha"
 /note="Vector: PORT2; Site_1: EcoRI; Site_2: XhoI; Sized
 fractionated cDNAs were directly ligated into PORT2.
 Plasmid cDNA library."
 BASE COUNT 158 a 106 c 124 g 110 t
 ORIGIN

Query Match 68.0%; Score 17; DB 10; Length 498;
 Best Local Similarity 100.0%; Pred. No. 5.3;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gcatgccacatttcg 17
 ||||||||||||||||

Db 266 GCATGCCACGATTTCG 250

RESULT 2
 A0030894/c 429 bp mRNA EST 29-OCT-1998
 LOCUS A0030894 Rice cDNA from immature leaf including apical meristem
 DEFINITION Oryza sativa cDNA clone E60405_1A, mRNA sequence.
 A0030894
 VERSION A0030894.1 GI:3766784
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Eriophytidae; Oryzaceae; Oryza.
 1 (bases 1 to 429)
 Sasaki, T. and Yamamoto, K.
 Rice cDNA from immature leaf including apical meristem
 Unpublished (1997)
 Contact: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
 305-8602, Japan
 Tel: 81-298-38-7441
 Fax: 81-298-38-7468
 Email: tsasaki@affrc.affrc.go.jp, URL: <http://rpg.dna.affrc.go.jp/>
 PROJECT "RGP".

FEATURES

source 1..429 Location/Qualifiers
 /organism="Oryza sativa"
 /strain="Nipponbare"
 /db_xref="taxon:4530"
 /clone="E60405_1A"
 /clone_lib="Rice cDNA from immature leaf including apical
 meristem"
 /dev_stage="Immature"
 /note="Organ: leaf; Immature leaf including apical
 meristem (under long day condition)"
 BASE COUNT 106 a 122 c 101 g 99 t 1 others
 ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 429;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 catgccacatttcg 17
 ||||||||||||||||
 Db 302 CATGCCACGATTTCG 287

RESULT 3
 A0031475/c 510 bp mRNA EST 30-OCT-1998
 LOCUS A0031475 Rice cDNA from immature leaf including apical meristem
 DEFINITION Oryza sativa cDNA clone E61659_2Z, mRNA sequence.
 A0031475
 VERSION A0031475.1 GI:3767365
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Eriophytidae; Oryzaceae; Oryza.
 1 (bases 1 to 510)
 Sasaki, T. and Yamamoto, K.
 Rice cDNA from immature leaf including apical meristem
 Unpublished (1997)
 Contact: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
 305-8602, Japan

Tel: 81-298-38-7441
 Fax: 81-298-38-7468
 Email: tsasaki@abr.affrc.go.jp, URL: http://irp.dna.affrc.go.jp/
 PROJECT = "RGP".

FEATURES

source

Location/Qualifiers

1. 510
 /organism="Oryza sativa"
 /strain="Nipponbare"
 /db_xref="taxon:4530"
 /clone="E61659_22"
 /clone.lib="Rice cDNA from immature leaf including apical meristem"
 /dev_stage="Immature"
 /note="Organ: leaf; immature leaf including apical meristem (under long day condition)"

BASE COUNT 176 a 98 c 99 g 137 t

Query Match 64.0%; Score 16; DB 10; Length 510;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagcattcg 17
 |||||

DB 136 CATGCCAGCATTTCG 121

RESULT 4

AU030895 666 bp mRNA EST 29-OCT-1998
 LOCUS AU030895/c
 DEFINITION Oryza sativa cDNA from immature leaf including apical meristem

ACCESSION AU030895
 VERSION AU030895.1 GI:3766785
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa.

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 666)
 Sasaki, T. and Yamamoto, K.

REFERENCE Rice cDNA from immature leaf including apical meristem
 TITLE Unpublished (1997)
 JOURNAL Contact: Takuji Sasaki
 COMMENT National Institute of Agrobiological Resources
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
 305-8602, Japan
 Tel: 81-298-38-7441
 Fax: 81-298-38-7468
 Email: tsasaki@abr.affrc.go.jp, URL: http://irp.dna.affrc.go.jp/
 PROJECT = "RGP".

FEATURES

source

Location/Qualifiers

1. 666
 /organism="Oryza sativa"
 /strain="Nipponbare"
 /db_xref="taxon:4530"
 /clone="E60405_22"
 /clone.lib="Rice cDNA from immature leaf including apical meristem"
 /dev_stage="Immature"
 /note="Organ: leaf; immature leaf including apical meristem (under long day condition)"

BASE COUNT 195 a 149 c 131 g 181 t 10 others

ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 666;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagcattcg 17

DB 293 CATGCCAGCATTTCG 278
 |||||

RESULT 5
 AU078750 704 bp mRNA EST 18-OCT-1999
 LOCUS AU078750/c
 DEFINITION E0723_42, mRNA sequence.

ACCESSION AU078750
 VERSION AU078750.1 GI:6062509
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa.

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 704)
 Sasaki, T. and Yamamoto, K.

REFERENCE Rice cDNA from panicle at flowering stage
 TITLE Unpublished (1996)
 JOURNAL Contact: Takuji Sasaki
 COMMENT National Institute of Agrobiological Resources
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
 305-8602, Japan
 Tel: 81-298-38-7441
 Fax: 81-298-38-7468
 Email: tsasaki@abr.affrc.go.jp, URL: http://irp.dna.affrc.go.jp/
 PROJECT = "RGP".

FEATURES

source

Location/Qualifiers

1. 704
 /organism="Oryza sativa"
 /strain="Nipponbare"
 /db_xref="taxon:4530"
 /clone="E0723_42"
 /clone.lib="Rice panicle at flowering stage"
 /dev_stage="flowering stage"
 /note="Organ: panicle; Rice cDNA from panicle at flowering stage"

BASE COUNT 203 a 167 c 142 g 191 t 1 others

ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 704;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagcattcg 17
 |||||

DB 346 CATGCCAGCATTTCG 331

RESULT 6
 AZ578088 258 bp DNA GSS 08-DEC-2000
 LOCUS AZ578088/c
 DEFINITION 18f02 Shot-gun genomic library of Rhizobium strain ANU265 Rhizobium

sp. NGR234 genomic clone 18f02, DNA sequence.
 ACCESSION AZ578088
 VERSION AZ578088.1 GI:11605015
 KEYWORDS GSS.

SOURCE Rhizobium sp. NGR234.
 ORGANISM Rhizobium sp. NGR234
 Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 Rhizobiaceae; Rhizobium.

REFERENCE 1 (bases 1 to 258)
 Virey, V., Rosenthal, A., Broughton, W.J. and Perret, X.
 TITLE Genetic snapshots of the Rhizobium species NGR234 genome
 JOURNAL genomebiology.com 1 (6), 0014.1-0014.7 (2000)
 COMMENT Contact: Virginie Virey
 Laboratoire de Biologie Moleculaire des Plantes Superieures
 University of Geneva
 1 Chemin de l'Imperatrice, Chambesey/Geneva 1292, Switzerland
 Tel: +44(0)1603450000

Fax: +44(0)1603450045
Email: virginie.viprey@bsrc.ac.uk
Class: shotgun.

FEATURES
source
1. .258
Location/Qualifiers
/organism="Rhizobium sp. NGR234"
/strain="AN0265"
/db_xref="taxon:394"
/clone_lib="Shot-gun genomic library of Rhizobium strain AN0265"
/note="Vector: M13; derivative strain of NGR234 cured of pNGR234a"

BASE COUNT
ORIGIN
51 a 82 c 79 g 46 t

Query Match
Best Local Similarity 100.0%; Pred. No. 71;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 tcgccagcatttcca 18
Db 175 TCGCCAGCATTTCGA 161

RESULT 7
BE065518 338 bp mRNA 09-JUN-2000
LOCUS RC3-BT0316-170200-014-b06 BT0316 Homo sapiens cDNA, mRNA sequence.
DEFINITION BE065518
VERSION BE065518.1 GI:8410168
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 338)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Britones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.P., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.G.

TITLE
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL
MEDLINE
COMMENT
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

FEATURES
source
1. .338
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0316"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles
High quality sequence start: 17
High quality sequence stop: 338.

into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT
ORIGIN
76 a 100 c 66 g 96 t

Query Match
Best Local Similarity 100.0%; Pred. No. 74;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 catgccagcatttc 16
Db 97 CATGCCAGCATTTCG 83

RESULT 8
AV551712 341 bp mRNA 06-SEP-2000
LOCUS AV551712 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
DEFINITION AV551712 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
CDNA clone R213605R 5', mRNA sequence.
ACCESSION AV551712
VERSION AV551712.1 GI:8723125
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana

REFERENCE
AUTHORS
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 341)
Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizukazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

JOURNAL
MEDLINE
COMMENT
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizukazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source
1. .341
Location/Qualifiers
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="R213605R"
/clone_lib="Arabidopsis thaliana roots Columbia"
/tissue_type="roots"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT
ORIGIN
73 a 85 c 53 g 130 t

Query Match
Best Local Similarity 100.0%; Pred. No. 74;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 tcgccagcatttcca 18
Db 322 TCGCCAGCATTTCGA 336

RESULT 9
H60965 393 bp mRNA 06-OCT-1995
LOCUS YR22607.r1 Soares fetal liver spleen 1NF5 Homo sapiens cDNA clone
DEFINITION IMAGE:206028 5', mRNA sequence.
ACCESSION H60965
VERSION H60965.1 GI:1013797
KEYWORDS EST.
SOURCE human.

[illegible]

TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 Insert Size: 1146
 High quality sequence stops: 180 Source: IMAGE Consortium, LLNL
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1146 Std Error: 0.00
 Seq primer: M13RPI
 High quality sequence stop: 180.
 FEATURES
 source
 1. 394
 /organism="Homo sapiens"
 /db_xref="GDB:468473"
 /db_xref="taxon:9606"
 /clone="IMAGE:112856"
 /clone_lib="Soares fetal liver spleen 1NFLS"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: pUT3D (Pharmacia)
 with a modified polylinker; Site.1: Pac I; Site.2: Eco RI;
 1st strand cDNA was primed with a Pac I...oligo(dT) primer
 15' AACGTGAGGATTATTAATTAAGATCTTTTTTTTTTTTTTTT 3'1,
 (pharmacia)-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pUT3 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M.Paulina Bonaldo."
 BASE COUNT 86 a 98 c 82 g 122 t 6 others
 ORIGIN
 Query Match 60.0%; Score 15; DB 11; Length 394;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 15; Conservative 0; Mismatches 0; Gaps 0
 Oy 2 catgccagcattc 16
 |||||||||||||
 Db 183 CATGCCAGCATTTTC 169
 RESULT 11
 LOCUS W23358 394 bp mRNA EST 12-DEC-1996
 DEFINITION SWAMCA507SK Brugia malayi adult male cDNA (SAM94ML-BMAN) Brugia
 W23358
 W23358 malayi cDNA clone SWAMCA507 5', mRNA sequence.
 ACCESSION W23358
 VERSION W23358.1 GI:1300212
 KEYWORDS EST.
 SOURCE Brugia malayi.
 ORGANISM Brugia malayi
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 Onchocercidae; Brugia.
 1 (bases 1 to 394)
 Williams, S.A.
 Genes expressed in adult males of Brugia malayi
 Unpublished (1995)
 Contact: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 413/853826
 Fax: 413/853786
 Email: genome@smith.edu
 Seq primer: pBluescript SK.
 Location/Qualifiers
 FEATURES
 CDS
 ORF1
 ORF2
 ORF3
 ORF4
 ORF5
 ORF6
 ORF7
 ORF8
 ORF9
 ORF10
 ORF11
 ORF12
 ORF13
 ORF14
 ORF15
 ORF16
 ORF17
 ORF18
 ORF19
 ORF20
 ORF21
 ORF22
 ORF23
 ORF24
 ORF25
 ORF26
 ORF27
 ORF28
 ORF29
 ORF30
 ORF31
 ORF32
 ORF33
 ORF34
 ORF35
 ORF36
 ORF37
 ORF38
 ORF39
 ORF40
 ORF41
 ORF42
 ORF43
 ORF44
 ORF45
 ORF46
 ORF47
 ORF48
 ORF49
 ORF50
 ORF51
 ORF52
 ORF53
 ORF54
 ORF55
 ORF56
 ORF57
 ORF58
 ORF59
 ORF60
 ORF61
 ORF62
 ORF63
 ORF64
 ORF65
 ORF66
 ORF67
 ORF68
 ORF69
 ORF70
 ORF71
 ORF72
 ORF73
 ORF74
 ORF75
 ORF76
 ORF77
 ORF78
 ORF79
 ORF80
 ORF81
 ORF82
 ORF83
 ORF84
 ORF85
 ORF86
 ORF87
 ORF88
 ORF89
 ORF90
 ORF91
 ORF92
 ORF93
 ORF94
 ORF95
 ORF96
 ORF97
 ORF98
 ORF99
 ORF100
 ORF101
 ORF102
 ORF103
 ORF104
 ORF105
 ORF106
 ORF107
 ORF108
 ORF109
 ORF110
 ORF111
 ORF112
 ORF113
 ORF114
 ORF115
 ORF116
 ORF117
 ORF118
 ORF119
 ORF120
 ORF121
 ORF122
 ORF123
 ORF124
 ORF125
 ORF126
 ORF127
 ORF128
 ORF129
 ORF130
 ORF131
 ORF132
 ORF133
 ORF134
 ORF135
 ORF136
 ORF137
 ORF138
 ORF139
 ORF140
 ORF141
 ORF142
 ORF143
 ORF144
 ORF145
 ORF146
 ORF147
 ORF148
 ORF149
 ORF150
 ORF151
 ORF152
 ORF153
 ORF154
 ORF155
 ORF156
 ORF157
 ORF158
 ORF159
 ORF160
 ORF161
 ORF162
 ORF163
 ORF164
 ORF165
 ORF166
 ORF167
 ORF168
 ORF169
 ORF170
 ORF171
 ORF172
 ORF173
 ORF174
 ORF175
 ORF176
 ORF177
 ORF178
 ORF179
 ORF180
 ORF181
 ORF182
 ORF183
 ORF184
 ORF185
 ORF186
 ORF187
 ORF188
 ORF189
 ORF190
 ORF191
 ORF192
 ORF193
 ORF194
 ORF195
 ORF196
 ORF197
 ORF198
 ORF199
 ORF200
 ORF201
 ORF202
 ORF203
 ORF204
 ORF205
 ORF206
 ORF207
 ORF208
 ORF209
 ORF210
 ORF211
 ORF212
 ORF213
 ORF214
 ORF215
 ORF216
 ORF217
 ORF218
 ORF219
 ORF220
 ORF221
 ORF222
 ORF223
 ORF224
 ORF225
 ORF226
 ORF227
 ORF228
 ORF229
 ORF230
 ORF231
 ORF232
 ORF233
 ORF234
 ORF235
 ORF236
 ORF237
 ORF238
 ORF239
 ORF240
 ORF241
 ORF242
 ORF243
 ORF244
 ORF245
 ORF246
 ORF247
 ORF248
 ORF249
 ORF250
 ORF251
 ORF252
 ORF253
 ORF254
 ORF255
 ORF256
 ORF257
 ORF258
 ORF259
 ORF260
 ORF261
 ORF262
 ORF263
 ORF264
 ORF265
 ORF266
 ORF267
 ORF268
 ORF269
 ORF270
 ORF271
 ORF272
 ORF273
 ORF274
 ORF275
 ORF276
 ORF277
 ORF278
 ORF279
 ORF280
 ORF281
 ORF282
 ORF283
 ORF284
 ORF285
 ORF286
 ORF287
 ORF288
 ORF289
 ORF290
 ORF291
 ORF292
 ORF293
 ORF294
 ORF295
 ORF296
 ORF297
 ORF298
 ORF299
 ORF300
 ORF301
 ORF302
 ORF303
 ORF304
 ORF305
 ORF306
 ORF307
 ORF308
 ORF309
 ORF310
 ORF311
 ORF312
 ORF313
 ORF314
 ORF315
 ORF316
 ORF317
 ORF318
 ORF319
 ORF320
 ORF321
 ORF322
 ORF323
 ORF324
 ORF325
 ORF326
 ORF327
 ORF328
 ORF329
 ORF330
 ORF331
 ORF332
 ORF333
 ORF334
 ORF335
 ORF336
 ORF337
 ORF338
 ORF339
 ORF340
 ORF341

```

source
1. 394
/organism="Brugia malayi"
/strain="TRs Labs"
/db_xref="taxon:6279"
/clone="SWAMCA507"
/lab_host="Brugia malayi adult male cDNA (SAM94NL-Bman)"
/note="Vector: lambda UniZap XR; Site_1: EcoR I; Site_2:
Xho I; Lymphatic filarial nematode parasite of humans.
mRNA was prepared from adult males of Brugia malayi
isolated from birds and converted to double stranded cDNA
using reverse transcriptase and oligo(dT) followed by
Rnase H and DNase I. The library had 4.6 x 10E6
independent recombinants and average insert size was 800
base pairs. The library was constructed by Noelle Ling.
The library is available from Dr. S.A. Williams, email
genome@smith.edu."

BASE COUNT      151 a      60 c      78 g      103 t      2 others
ORIGIN

Query Match      60.0%; Score 15; DB 11; Length 394;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 gccacgattcgcgc 20
Db      32 GCCACGATTTCGATC 46

RESULT 12
BE044431      399 bp      mRNA      EST      08-JUN-2000
LOCUS      ho45604.x1 Soares_NFL.T.GBC.S1 Homo sapiens cDNA clone
DEFINITION      IMAGE:3040327 3', mRNA sequence.
ACCESSION      BE044431
VERSION      BE044431.1 GI:8361484
KEYWORDS      human.
SOURCE      Homo sapiens
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 399)
REFERENCE      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE      Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seg primer: -400p from Gibco.
Location/Qualifiers
1. 399
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3040327"
/clone_lib="Soares_NFL.T.GBC.S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCL-CGAP-GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo. "

BASE COUNT      91 a      94 c      96 g      118 t
ORIGIN

Query Match      60.0%; Score 15; DB 10; Length 399;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 catgcgcgatttc 16
Db      329 CATGCCGACGATTTC 343

RESULT 13
AT268648      414 bp      mRNA      EST      17-NOV-1998
LOCUS      q039f05.x1 NCI-CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1910913 3',
DEFINITION      mRNA sequence.
ACCESSION      AT268648
VERSION      AT268648.1 GI:3887815
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 414)
REFERENCE      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE      Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seg primer: -400p from Gibco
High quality sequence stop: 398.
Location/Qualifiers
1. 414
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1910913"
/clone_lib="NCI-CGAP_Lu5"
/lisue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
neuroendocrine lung carcinoid, and was then primed with a
Not I - oligo(dT) primer. Double-stranded cDNA was ligated
to Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library is normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT      122 a      82 c      109 g      101 t
ORIGIN

Query Match      60.0%; Score 15; DB 10; Length 414;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 catgcgcgatttc 16
Db      353 CATGCCGACGATTTC 367

RESULT 14
AO027122      435 bp      DNA      GSS      30-JUN-1998
LOCUS      CIT-HSP-2323H13.TF CIT-HSP Homo sapiens genomic clone 2323H13, DNA
DEFINITION

```

ACCESSION	sequence.
VERSION	A0027122
KEYWORDS	A0027122.1 GI:3267344
SOURCE	GSS.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE	1 (bases 1 to 435)
JOURNAL	Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.
COMMENT	Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)
	Unpublished (1998)
	Contact: Mark Adams
	Department of Eukaryotic Genomics
	The Institute for Genomic Research
	9712 Medical Center Dr., Rockville, MD 20850, USA
	Tel: 301 838 0200
	Fax: 301 838 0208
	Email: mdadams@tigr.org
	Clones are available from Research Genetics (info@resgen.com). BAC
	End Search page:
	http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
	Seq primer: M13-21
	Class: BAC ends.
FEATURES	Location/Qualifiers
SOURCE	1..435
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="2323H13"
	/clone_lib="CIT-HSP"
	/sex="Male"
	/cell_type="Sperm"
	/note="Vector: pheloBAC11; Site_1: HindIII; Site_2: HindIII"
BASE COUNT	140 a 97 c 77 g 121 t
ORIGIN	
Query Match	60.0%; Score 15; DB 13; Length 435;
Best Local Similarity	100.0%; Pred. No. 76;
Matches 15; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY 2 catgccagcagcattc 16	
Db 178 CATGCCAGCAGCATTTC 192	
RESULT 15	
H69200/c	437 bp mRNA EST 19-OCT-1995
LOCUS	yr95f10.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
DEFINITION	IMAGE:213067 5', mRNA sequence.
ACCESSION	H69200
VERSION	H69200.1 GI:1030526
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (bases 1 to 437)
	Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chapelli,B., Chissoe,S., Dietrich,N., Dubucq,T., Favello,A., Gish,M., Hawkins,M., Hultman,M., Kucaba,T., Lacey,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Ritzkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevisan,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
TITLE	Generation and analysis of 280,000 human expressed sequence tags
JOURNAL	Genome Res. 6 (9), 807-828 (1996)
MEDLINE	97044478
COMMENT	Contact: Wilson RK

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Insert Size: 1098
High quality sequence stops: 300
Source: IMAGE Consortium, LNL.
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lln.gov) for further information.
Insert Length: 1098 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 300.
Location/Qualifiers
1..437
/organism="Homo sapiens"
/db_xref="GDB:3778371"
/db_xref="taxon:9606"
/clone="IMAGE:213067"
/clone_lib="Soares fetal Liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pTR73D (Pharmacia)
with a modified polylinker; Site.1: Pac I; Site.2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACCTGCAAGCAATTATTAAGATCTTTTCTTTTCTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pTR73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaudo."

RESULT	16		
LOCUS	N92081/c		
DEFINITION	N92081	438 bp	MRNA EST
ACCESSION	2a23h03.r1	Soares fetal liver spleen	INFLS Homo sapiens
VERSION	IMAGE:293429.5	*, mRNA sequence.	04-APR-1996
KEYWORDS	N92081		
SOURCE	N92081.1	GI:1264390	
ORGANISM	EST.		
	human.		
	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
REFERENCE	1	(bases 1 to 438)	
AUTHORS	Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman		
	M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Maitra,M., Parsons,J.		
	Riffin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Watson		
	,R., Williamson,A., Wohldmann,P. and Wilson,R.		
TITLE	The Washu-Merck EST project		
JOURNAL	Unpublished (1995)		
COMMENT	Contact: Wilson KK		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@watson.wustl.edu		
	This clone is available royalty-free through LNL; contact the		
	IMAGE consortium (info@image.lnl.gov) for further information.		
	Seq primer: mob.RBGA-ET		

FEATURES High quality sequence stop: 321.

Source location/Qualifiers
1. 438

/organism="Homo sapiens"
/db_xref="GDB:3801206"
/db_xref="taxon:9606"
/clone="IMAGE:293429"
/clone_1ib="Soares fetal liver spleen 1NPLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I oligo(dT) primer [5' AACGCGACGACATTTATTAAGATCTTTTCTTTTCTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 104 a 111 c 101 g 118 t 4 others
ORIGIN

Query Match
Best local Similarity 100.0%; Score 15; DB 11; Length 438;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagcatttc 16
Db 308 CATGCCAGCATTTTC 294

RESULT 17
AV722315/c 445 bp mRNA EST 16-OCT-2000
LOCUS AV722315 HTB Homo sapiens cDNA clone HTBAMG01 5', mRNA sequence.
DEFINITION AV722315
ACCESSION AV722315
VERSION AV722315.1 GI:10824678
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 445)
AUTHORS Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R., Chen, J., Chen, Z., and Han, Z.
Homo sapiens cDNA HTB clones
Unpublished (2000)
Contact: Zenguan Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
Tel: 86-21-50801919 (ex. 45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES Location/Qualifiers
1. 445

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTBAMG01"
/clone_1ib="HTB"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 102 a 123 c 104 g 116 t
ORIGIN

Query Match
Best local Similarity 100.0%; Score 15; DB 10; Length 445;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagcatttc 16
Db 296 CATGCCAGCATTTTC 282

RESULT 18
AA026650/c 466 bp mRNA EST 09-MAY-1997
LOCUS zj99e08.1 Soares-pregnant_uterus_NbHPU Homo sapiens cDNA clone
DEFINITION IMAGE:469190 5', mRNA sequence.
ACCESSION AA026650
VERSION AA026650.1 GI:1492715
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 466)
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapelli, B., Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through INM; contact the IMAGE Consortium (info@image.jnl.gov) for further information.
Insert Length: 1129 Std Error: 0.00
Seq primer: -28M13 rev2 from Amerisham
High quality sequence stop: 364.

FEATURES Location/Qualifiers
1. 466

/organism="Homo sapiens"
/db_xref="GDB:3755192"
/db_xref="taxon:9606"
/clone="IMAGE:469190"
/clone_1ib="Soares-pregnant_uterus_NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT73-Pac; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACGCGACGACATTTGCGGCGCTTTTCTTTTCTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."

BASE COUNT 114 a 124 c 100 g 128 t
ORIGIN

Query Match
Best local Similarity 100.0%; Score 15; DB 10; Length 466;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagcatttc 16
Db 263 CATGCCAGCATTTTC 249

RESULTS	19
H60979/c	
LOCUS	H60979 475 bp mRNA EST 06-OCT-1995
DEFINITION	y122f01.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone IMAGE:206041 5', mRNA sequence.
ACCESSION	H60979
VERSION	H60979
KEYWORDS	GI:1013811
SOURCE	EST.
ORGANISM	human.
TITLE	Homo sapiens
JOURNAL	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 475)
AUTHORS	Hallier,L., Clark,N., Dubque,T., Elliston,K., Hawkins,M., Holman ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,D., Rikkin,L., Rohlfing,T., Soares,M., Tan,F., Teyvaskis,E., Waterston ,R., Williamson,A., Wohlmann,P. and Wilson,R. The Washu-Merck Est Project Unpublished (1995) Contact: Wilson RK Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Insert Size: 1101 High quality sequence stops: 339 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL : contact the IMAGE Consortium (infoimage.llnl.gov) for further information. Insert Length: 1101 Std Error: 0.00 Seq primer: M13RP1 High quality sequence stop: 339. Location/Qualifiers 1..475 /organism="Homo sapiens" /db_xref="GDB:3775172" /db_xref="taxon:9606" /clone="IMAGE:206041" /clone_lib="Soares fetal liver spleen INFLS" /sex="male" /dev_stage="20 week-post conception fetus" /lab_host="DH10B (ampicillin resistant)" /note="organ: Liver and Spleen; Vector: pT73D (Pharmacia)a with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACGGAAGAATTAATTAAAGATCGTTTGTGGTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Patima Bonaído."
FEATURES	
source	
BASE COUNT	109 a 123 c 95 g 143 t 5 others
ORIGIN	
Query Match	60.0%; Score 15; DB 11; Length 475;
Best Local Similarity	100.0%, Pred. No. 77;
Matches 15; Conservative	0; Mismatches 0; Indels 0; Gaps 0
OY	2 catgccacgacattcc 16
Db	119 CATGCCACGACATTTC 105
RESULT 20	
AI692369	
LOCUS	AI692369 491 bp mRNA EST 17-DEC-1999
DEFINITION	wd63102.x1 NCL_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2336283 3', mRNA sequence.
ACCESSION	AI692369

VERSION	EST.	AI692369.1	GI:4969709
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
ATTNORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 491)		
	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
JOURNAL	Tumor Gene Index		
COMMENT	Unpublished (1997)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgaps-rt@mail.nih.gov		
	Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.		
	Emmert-Buck, M.D., Ph.D.		
	CDNA Library Preparation: M. Bento Soares, Ph.D.		
	CDNA Library Arrayed by: Greg Lennon, Ph.D.		
	DNA Sequencing by: Washington University Genome Sequencing Center		
	Clone distribution: NCI-CGAP clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	www-bio.llnl.gov/bdpr/image/image.html		
	Insert Length: 589 Std Error: 0.00		
	Seq primer: -40UP from GldCP		
FEATURES	High quality sequence stop: 471.		
Source	location/Qualifiers		
	1..491		

```

BASE COUNT      151 a      94 c      129 g      116 t      1 others
ORIGIN

Query Match      60.0%; Score 15; DB 10; Length 491;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2      catgccacacatttc 16
        |||
Db      344      CATGCCACACATTTC 358

RESULT 21
BF211097/c
LOCUS      BF211097
DEFINITION      BF211097      491 bp      mRNA
                6018127106F1 NIH_MGC_54 Homo sapiens cDNA clone IMAGE:4047120 5',
                mRNA sequence.
ACCESSION      BF211097
VERSION      BF211097
KEYWORDS      BF211097.1 GI:11104683
SOURCE      EST.
            human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 491)
REFERENCE      NIH-MGC http://mhc.nci.nih.gov/.
AUTHORS      National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE      Unpublished (1999)
JOURNAL      Contact: Robert Strausberg, Ph.D.
COMMENT      Email: cgabs-remail.nih.gov
              Tissue Procurement: ATCC

```

CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LCM875 row: 1 column: 01
 High quality sequence start: 56
 High quality sequence stop: 491.
 Location/Qualifiers
 1. 491

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4047120"
 /clone_1lb="NIH_MGC_54"
 /issue_type="from chronic myelogenous leukemia"
 /lab_host="DH10B (v1 phage-resistant)"
 /note="Organ: bone marrow; Vector: pDNR-11B (Clontech);
 Site: 1: SfiI (ggccgctcgccg); Site: 2: SfiI (ggccatagctc
); Double-stranded cDNA was prepared from cell line RNA.
 5' and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGGCCATTATGACC-3' and 3' adaptor
 sequence: 5'-ATTCTAGAGCGCGAGCGCGACATG-dT(30)BN-3'
 (where B = A, C, G, or T and N = A, C, G, or T). Average
 insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA)."
 BASE COUNT 123 a 128 c 119 g 121 t
 ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 491;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 catcgccagcatctc 16
 |||||||||||||
 Db 305 CATCGCCAGCATTTTC 291

RESULT 22
 A2924681 492 bp DNA GSS 01-APR-2001
 LOCUS 4906.id55a24.s1 Saccharomyces mikatae IPO 1815 Saccharomyces
 DEFINITION mikatae genomic clone 4906.id55a24.s1, DNA sequence.
 ACCESSION A2924681
 VERSION A2924681.1 GI:13495580
 KEYWORDS GSS.
 SOURCE Saccharomyces mikatae.
 ORGANISM Saccharomyces mikatae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 TITLE 1 (bases 1 to 492)
 AUTHORS Cliffter,P.F., Hillier,L.W., Pilton,L., Graves,T., Miner,T., Gish
 S., W.R., Waterston,R.H. and Johnston,M.
 JOURNAL Surveying Saccharomyces genomes to identify functional elements by
 comparative DNA sequence analysis
 COMMENT Unpublished (2001)
 CONTACT: Johnston M
 Department of Genetics
 Washington University Medical School
 Box 8332, 4566 Scott Ave., St. Louis, MO 63110, USA
 Tel: 314 362 2735
 Fax: 314 362 7855
 Email: mjgenetics.wustl.edu
 Class: random plasmid subclone.
 Location/Qualifiers
 1. 492
 /organism="Saccharomyces mikatae"
 /strain="IFO 1815"
 /db_xref="taxon:114525"
 /clone="4906.id55a24.s1"

/clone_1lb="Saccharomyces mikatae IPO 1815"
 /note="Random genomic sequence"
 BASE COUNT 149 a 83 c 91 g 169 t
 ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 492;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 ccagcatcgatcc 21
 |||||||||||||
 Db 475 CCAGCATTCGATCC 461

RESULT 23
 T84279 507 bp mRNA EST 16-MAR-1995
 LOCUS Y847B04.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
 DEFINITION IMAGE:111343 5', mRNA sequence.
 ACCESSION T84279
 VERSION T84279.1 GI:712567
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 TITLE 1 (bases 1 to 507)
 JOURNAL Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 COMMENT The Washu-Merck EST Project
 Unpublished (1995)
 CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 Insert Size: 1225
 High quality sequence stops: 283 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert length: 1225 Std Error: 0.00
 Seq primer: M13RP1
 High quality sequence stop: 283.
 Location/Qualifiers
 1. 507
 /organism="Homo sapiens"
 /db_xref="GDB:466960"
 /db_xref="taxon:9606"
 /clone="IMAGE:111343"
 /clone_1lb="Soares fetal liver spleen INFLS"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
 with a modified polylinker; Site: 1: Pac I; Site: 2: Eco RI;
 1st strand cDNA was primed with a Pac I - oligo(dT) primer
 5' AACGCGAGATTAATTAAGATCTTTTTTTTTTTT 3',
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."

FEATURES

source
 /organism="Homo sapiens"
 /db_xref="GDB:466960"
 /db_xref="taxon:9606"
 /clone="IMAGE:111343"
 /clone_1lb="Soares fetal liver spleen INFLS"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
 with a modified polylinker; Site: 1: Pac I; Site: 2: Eco RI;
 1st strand cDNA was primed with a Pac I - oligo(dT) primer
 5' AACGCGAGATTAATTAAGATCTTTTTTTTTTTT 3',
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."
 BASE COUNT 106 a 125 c 114 g 159 t
 ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 507;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgccagcatttc 16
 Db 192 CATGCCAGCATTTC 178

RESULT 24

LOCUS H74120 528 bp mRNA EST 31-OCT-1995
 DEFINITION y515e08.f1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
 IMAGE:214886 5', mRNA sequence.

ACCESSION H74120
 VERSION H74120.1 GI:1047332

KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 528)
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, R.,
 Chissee, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins,
 M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Mardis, E., Moore,
 B., Morris, M., Parsons, J., Prange, C., Rifkin, B., Rohlfing, T.,
 Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,
 Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)

REFERENCE AUTHORS

TITLE JOURNAL
 MEDLINE
 COMMENT
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: eslewatson.wustl.edu
 Insert Size: 1036
 High quality sequence stops: 311
 Source: IMAGE Consortium, LLNL
 This clone is available royalty-free through LLNL: contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1036 Std Error: 0.00
 Seq primer: M13P1
 High quality sequence stop: 311.
 Location/Qualifiers

FEATURES

source
 1..528
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:214886"
 /clone_lib="Soares fetal liver spleen INFLS"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="PH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
 with a modified polylinker; Site_1: Pac I - oligo(dT) primer
 1st strand cDNA was primed with a Pac I - oligo(dT) primer
 15' AACGCGAGATTAATTAAGATCTTTTCTTTTCTTTT 3';
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."
 BASE COUNT 121 a 124 c 110 g 166 t 7 others

Query Match 60.0%; Score 15; DB 11; Length 528;
 Best Local Similarity 100.0%; Pred. No. 78;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgccagcatttc 16
 Db 107 CATGCCAGCATTTC 93

RESULT 25
 LOCUS BF839030 539 bp mRNA EST 13-JAN-2001
 DEFINITION PM2-HT0353-271100-011-h09 HT0353 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF839030
 VERSION BF839030.1 GI:12191373

KEYWORDS SOURCE ORGANISM

human.
 Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 539)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, M. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE JOURNAL MEDLINE COMMENT

Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM2&t2=PM2-HT0353-271100-011-h09&t3=2000-11-27&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence stop: 8
 Location/Qualifiers

FEATURES

source
 1..539
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HT0353"
 /dev_stage="Adult"
 /note="Organ: head, neck; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 124 a 143 c 128 g 143 t 1 others

Query Match 60.0%; Score 15; DB 11; Length 539;
 Best Local Similarity 100.0%; Pred. No. 78;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgccagcatttc 16
 Db 400 CATGCCAGCATTTC 386

RESULT 26

LOCUS AM588900 547 bp mRNA EST 10-MAY-2001
 DEFINITION rat06905.y2 Bird-Rao Meloidogyne incognita J2 Meloidogyne incognita
 cDNA 5' similar to TR:055013 055013 BERT3. contains Alu repetitive
 element; mRNA sequence.

ACCESSION AM588900
 VERSION AM588900.1 GI:7275932
 KEYWORDS EST.
 SOURCE Southern root-knot nematode.

Sequencing Center clone distribution: Research Genetics web
address: <http://www.researchgenetics.com/>

Seq primer: T3 ET from AmerSham
High quality sequence stop: 390.

Location/Qualifiers

FEATURES

source

1..622
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="Zebrafish Research Genetics C32 fin"
/tissue_type="Fin"
/lab_host="GeneHogs (HS996, a phage-resistant isolate of DH10B)"
/note="Vector: pT73b-Pac with a modified polylinker:
Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was prepared
from zebrafish(C32) fin, and was then primed with a Not I
- oligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library is non-normalized. Library was
constructed by Ning Wu. NOTE: Clones from this library are
only available thru Research Genetics (www.resgen.com)."

BASE COUNT

164 a 160 c 164 g 134 t

Query Match 60.0%; Score 15; DB 10; Length 622;
Best Local Similarity 100.0%; Pred. No. 79;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 tcgcgcagcttcga 18
|||||

Db 241 TCGCCAGCATTTGCA 227

RESULT 29
BE583417/c 630 bp mRNA EST 16-AUG-2000
LOCUS
DEFINITION
11-6H-MY Psojaemy Phytophthora sojae cDNA, mRNA sequence.
ACCESSION
BE583417 GI:9834366
VERSION
KEYWORDS
EST.
Phytophthora sojae.
SOURCE
Phytophthora sojae
ORGANISM
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora.

REFERENCE
1 (bases 1 to 630)
Qutob,D., Hrabert,P.T., Sobral,B.W.S. and Gijzen,M.
Comparative analysis of expressed sequences in Phytophthora sojae
JOURNAL
Plant Physiol. 123 (1), 243-254 (2000)
MEDLINE
20267956

COMMENT
Contact: Gijzen M
Agriculture and Agri-Food Canada
1391 Sandford Street, London, Ontario, Canada N5V 4T3
Tel: 519 457 1470
Fax: 519 457 3997
Email: gijzenm@agr.ca.

FEATURES

source

1..630
/organism="Phytophthora sojae"
/strain="race 2, strain p6497"
/db_xref="taxon:67593"
/clone_lib="Psojaemy"
/dev_stage="mycelium"
/lab_host="E. coli strain XL0R"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI. This
cDNA library was constructed from polyA+ enriched mRNA
from mycelium grown in liquid medium for 3 weeks at 25 C
in the dark in a liquid synthetic medium of 2.4 g sucrose
and 0.27 g asparagine, 10 mg cholesterol, 10 mg ascorbate,
and 2 mg thiamine, per liter, plus salts and minerals.
Complementary DNA was synthesized from mRNA using an
XhoI-poly(dT) linker-primer. EcoRI adaptors were ligated
to the blunt-ended cDNA fragments and the products were

digested with XhoI for directional cloning into lambda ZAP
Express vector. This lambda library was amplified once
using E. coli host strain XL1 Blue MRF+. Inserts were then
subcloned by mass excision using EXAssist helper phage for
conversion into phagemid vector pBK-CMV in E. coli host
strain XL0R. Sequenced using T3 primer: 5' ATT AAC CCG
CAC TAA AGG GA 3'.

BASE COUNT

135 a 184 c 169 g 140 t 2 others

Query Match 60.0%; Score 15; DB 10; Length 630;
Best Local Similarity 100.0%; Pred. No. 79;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgcgcagcttc 16
|||||

Db 16 CATGCCAGCATTTTC 2

RESULT 30
BH040083/c 643 bp DNA GSS 17-JUL-2001
LOCUS
DEFINITION
PC1-24-284M12.TV PC1-24 Mus musculus genomic clone PC1-24-284M12
, DNA sequence.
ACCESSION
BH040083
BH040083.1 GI:14818692
VERSION
KEYWORDS
GSS.
SOURCE
house mouse.
MUS musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murine; Mus.
1 (bases 1 to 643)
REFERENCE
Zhao S., Nierman,W., Malek,J., Shatsman,S., Akincel,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shwartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library PC1-24
Unpublished (1999)
Other GSSs: PC1-24-284M12.TJ
COMMENT
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library PC1-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.html>). BAC end
plate: 284 row: M column: 12
Seq primer: 47
Class: BAC ends.

FEATURES

source

1..643
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="PC1-24-284M12"
/clone_lib="PC1-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTAKBAC1; Site_1: BamHI; Site_2: BamHI.
PC1-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTAKBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT

183 a 156 c 175 g 129 t

ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 643;

Best Local Similarity 100.0%; Pred. No. 79;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cagcattcgatcct 22
|||||
Db 45 CAGCATTTCGATTCCT 31

RESULT 31
BF156891/c 684 bp mRNA EST 23-FEB-2001
LOCUS
DEFINITION
t156812.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
3618062 5' similar to TR:Q9UJ41 Q9UJ41 RAB5 GDP/GTP EXCHANGE FACTOR
HOMOLOGUE. ; mRNA sequence.

ACCESSION
BF156891
VERSION
KEYWORDS
SOURCE
ORGANISM
EST.
Zebrafish.
Danio rerio

REFERENCE
AUTHORS
Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy
, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood
, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,
Swaller, T., Gibbons, M., Pape, D., Harey, N., Schurk, R., Ritter, E.,
Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, K.
and Wilson, R.
WashU zebrafish EST Project 1998
Unpublished (1998)

TITLE
JOURNAL
COMMENT
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: zbrafish@wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center Clone
distribution information can be found through the I.M.A.G.E.
Consortium/ILMI, send email to: info@image.llnl.gov
Seq primer: r3 ET from Amersham
High quality sequence stop: 501.

FEATURES
source
Location/Qualifiers

1..684
/organism="Danio rerio"
/strain="Ab"
/db_xref="taxon:7955"
/clone="3818062"
/clone_lib="Sugano Kawakami zebrafish DRA"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/note="Vector: pME18-FL3; Site_1: DraIII (CACGCTG);
Site_2: DraIII (CACGCTG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGtggcctttttttttttttttt];
double-stranded cDNA was ligated to a DraIII adaptor
[tgggtgacctgg], digested and cloned into distinct DraIII
sites of the pME18-FL3 vector (5' site CACGCTG, 3' site
CACGCTG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and
kindly donated by Dr. Koichi Kawakami. Custom primers for
sequencing: 5' end primer CTTCCTCTCTAAAGCTGCG and 3' end
primer CGACCTGACCTGACGACA."

BASE COUNT 177 a 177 c 193 g 137 t
ORIGIN
Query Match 60.0%; Score 15; DB 11; Length 684;
Best local Similarity 100.0%; Pred. No. 80;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcgccagcattcgca 18
|||||
Db 237 TCGCCAGCATTTGCA 223

RESULT 32
BH039916/c 776 bp DNA GSS 17-JUL-2001
LOCUS
DEFINITION
RPCT-24-284K20.TV RPCT-24 Mus musculus genomic clone RPCT-24-284K20
DNA sequence.

ACCESSION
BH039916
VERSION
KEYWORDS
SOURCE
ORGANISM
GSS.
house mouse.
Mus musculus

REFERENCE
AUTHORS
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
1 (bases 1 to 776)
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akiret, B., Levins, M.,
Tsegaye, G., Geer, K., Krol, M., Shwartsbeyn, A., Gebregeorgis, E.,
Russell, D., de Jong, P. and Fraser, C.M.
Mouse BAC End Sequences from Library RPCT-24
Unpublished (1999)
Other GSSs: RPCT-24-284K20.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCT-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@gmail.com). Clones may be purchased from BACPAC
Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 264 row: K column: 20
Seq primer: T7
Class: BAC ends.

FEATURES
source
Location/Qualifiers

1..776
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCT-24-284K20"
/clone_lib="RPCT-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pPARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCT-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pPARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT 209 a 200 c 207 g 160 t
ORIGIN
Query Match 60.0%; Score 15; DB 13; Length 776;
Best local Similarity 100.0%; Pred. No. 81;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cagcattcgatcct 22
|||||
Db 44 CAGCATTTCGATTCCT 30

RESULT 33
BII13864/c 874 bp mRNA EST 26-JUN-2001
LOCUS
DEFINITION
602860705F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:5020104 5',
mRNA sequence.

ACCESSION B1113864
 VERSION B1113864.1 GI:14564765
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 874)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LCM1836 row: 1 column: 01
 High quality sequence stop: 599.
 Location/Qualifiers
 1..874
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5020104"
 /clone_lib="NIH-MGC_17"
 /tissue_type="Pharyngosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: muscle; Vector: pORF7; Site:1: EcoRI; Site:2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(6). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 250 a 212 c 236 g 176 t
 ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 874;
 Best Local Similarity 100.0%; Pred. No. 82;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 cagcattcgatcct 22
 ||||||||||||
 Db 857 CAGCATTCGATCCT 843

RESULT 34
 BG758892 986 bp mRNA EST 15-MAY-2001
 LOCUS 602713004F1 NIH-MGC_48 Homo sapiens cDNA clone IMAGE:4853400 5',
 DEFINITION mRNA sequence.
 ACCESSION BG758892
 VERSION BG758892.1 GI:14069545
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 986)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LCM1699 row: g column: 01
 High quality sequence stop: 824.
 Location/Qualifiers
 1..986
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4853400"
 /clone_lib="NIH-MGC_48"
 /tissue_type="primary B-cells from tonsils (cell line)"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: B-cells; Vector: pORF7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(6). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC library."

BASE COUNT 226 a 240 c 236 g 284 t
 ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 986;
 Best Local Similarity 100.0%; Pred. No. 83;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 catccgacgatttc 16
 ||||||||||||
 Db 565 CATCCGACGATTTC 551

RESULT 35
 BF144032 1096 bp mRNA EST 24-OCT-2000
 LOCUS 601791288P1 NCI-CGAP_Lu30 Mus musculus cDNA clone IMAGE:4022252 5',
 DEFINITION mRNA sequence.
 ACCESSION BF144032
 VERSION BF144032.1 GI:10983072
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1096)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LLM9278 row: o column: 21
 High quality sequence stop: 630.
 Location/Qualifiers
 1..1096
 /organism="Mus musculus"
 /strain="CZECH IT"
 /db_xref="taxon:10090"
 /clone="IMAGE:4022252"
 /clone_lib="NCI-CGAP_Lu30"
 /tissue_type="tumor, metastatic to mammary"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: SalI; transgenic model WNT-1, expression driven by

MTV-LTR enhancer: Cloned unidirectionally. Primer: Oligo
dT. Library constructed by Life Technologies.
Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 336 a 271 c 271 g 218 t

Query Match 60.0%; Score 15; DB 11; Length 1096;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 gcatttcgaccccg 24
|||||

Db 1076 GCATTTCGATCCTCG 1062

RESULT 36

LOCUS W06020 195 bp mRNA EST 22-MAY-2000
DEFINITION T8ESTy87f10.r1 TgRH Tachyzoite cDNA Toxoplasma gondii cDNA clone
tgy87f10.r1 5', mRNA sequence.

ACCESSION W06020
VERSION W06020.1 GI:1278733

KEYWORDS EST.
SOURCE Toxoplasma gondii.
ORGANISM Toxoplasma gondii

REFERENCE Eukaryota: Alveolata: Apicomplexa: Coccidia: Eimeriidae;
Sarcocystidae: Toxoplasma.
1 (bases 1 to 195)

AUTHORS

Hehl, A., Manger, I., Marra, M., Sibley, L.D., Ajioke, J.A., Aslett, M.A.,
Dietrich, N., Duboue, T., Hillier, L., Kucaba, T., Wan, K.L.,
Waterston, R.H. and Boothroyd, J.

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxowatson.wustl.edu

COMMENT

Unpublished (1996)
Contact: Marra M
Washu-Merck EST Project

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxowatson.wustl.edu

David Sibley at toxowatson.wustl.edu for further information
relating to organism, clone or library availability.

Seq primer: T3
High quality sequence stop: 148.

FEATURES

source location/Qualifiers
1..195
/organism="Toxoplasma gondii"
/strain="RH"
/db_xref="taxon:5811"
/clone="tgy87f10.r1"
/lab_host="X11-Blue MRF"
/note="Vector: Lambda ZAP. Site 1: EcoRI. Site 2: XhoI;
Toxoplasma RH strain tachyzoites were grown in human
foreskin fibroblast cultures in vitro. The library was
constructed by K.L. Wan, Cambridge University. cDNAs were
synthesized from polyA RNAs by oligo d(T) priming and
directionally cloned into the EcoRI to XhoI sites of the
Lambda Zapit vector using the ZAP-cDNA synthesis kit
(Stratagene). WARNING: the library contains a small
percentage of cDNAs derived from the human host cells."

BASE COUNT 59 a 45 c 62 g 27 t
ORIGIN

Query Match 56.0%; Score 14; DB 11; Length 195;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 gcatttcgaccccg 23
|||||

Db 181 GCATTTCGATCCTCG 168

RESULT 37

LOCUS A2922752 200 bp DNA GSS 22-MAR-2001
DEFINITION SLCot4A05 Sorghum bicolor SLCot Sorghum bicolor genomic DNA
sequence.

ACCESSION A2922752
VERSION A2922752.1 GI:13432973

KEYWORDS GSS.

SOURCE Sorghum.

ORGANISM Sorghum bicolor

REFERENCE Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 200)

AUTHORS Peterson, D.G., Schuizer, S.R., Lee, S.A., Sclara, E.B., Nagel, A.,
Tibbitts, D.C., Wessler, S.R. and Peterson, A.H.

Characterization of the Sorghum bicolor genome using DNA
renaturation kinetics (Cot analysis) and repetition-based cloning
Unpublished (2001)

COMMENT

Plant Genome Mapping Laboratory
University of Georgia
Room 162, Riverbend Research Bldg., 110 Riverbend Rd., Athens, GA
30602, USA
Tel: 706-583-0167
Fax: 706-583-0160
Email: dgp@arches.uga.edu

Class: Hydroxyapatite-fractionated DNA.

location/Qualifiers
1..200

FEATURES

source

/organism="Sorghum bicolor"

/cultivar="Brix623"

/db_xref="taxon:4558"

/clone_lib="Sorghum bicolor SLCot"

/lisse_type="leaves"

/dev_stage="seedling"

/note="Vector: pGEM-TA-Easy; A Cot analysis was performed
for the sorghum genome. Based on the resulting Cot curve,
hydroxyapatite chromatography was used to isolate
'highly-repetitive' (HR), 'moderately-repetitive' (MR),
and 'single/low-copy' (SL) sequence components from
sheared genomic DNA. The three repetition-based DNA
components were cloned into E. coli to produce HRCot,
MRCot, and SLCot genomic libraries. Blotting and
sequencing data indicates that each library is
representative of the component from which it was derived.
putative ID listings given for sequences are based on
comparison (blastn) with sequences in the NCBI Nr
database. Only the primary match is given (all primary E
values are < or =3d1.00E-5). In no instance does a 'Cot
clone' contain the complete sequence of its putative Nr
match."

BASE COUNT 48 a 49 c 57 g 42 t 4 others
ORIGIN

Query Match 56.0%; Score 14; DB 13; Length 200;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catcgccagacatt 15
|||||

Db 91 CATCGCCAGCATTT 104

RESULT 38
LOCUS AW065325/c 203 bp mRNA EST 12-OCT-1999
DEFINITION 614045H11.y1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
mRNA sequence.

ACCESSION AW065325

```

VERSION      AM065325.1  GI:6020397
KEYWORDS     EST.
SOURCE       Zea mays.
ORGANISM     Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
              clade; Panicoideae; Andropogoneae; Zea.
REFERENCE    1 (bases 1 to 203)
AUTHORS      Walbot,V.
TITLE        Maize ESTs from various cDNA libraries sequenced at Stanford
              University
JOURNAL      Unpublished (1999)
COMMENT      Contact: Walbot V
              Department of Biological Sciences
              Stanford University
              855 California Ave, Palo Alto, CA 94304, USA
              Tel: 650 723 2227
              Fax: 650 725 8221
              Email: walbot@stanford.edu
              Plate: 614045 row: H column: 11.
FEATURES     Location/Qualifiers
              1..203
              /organism="Zea mays"
              /cultivar="W23"
              /db_xref="taxon:4577"
              /clone_lib="G14 - root cDNA library from Walbot Lab"
              /tissue_type="root"
              /dev_stage="3-4 days old"
              /lab_host="X10LR"
              /note="Organ: root; Vector: pBluescriptII SK+; Site_1:
              EcORI; Site_2: XhoI; 3-4 days old root tissue from Walbot
              Lab (LM)"
BASE COUNT   46 a 52 c 51 g 54 t
ORIGIN
Query Match 56.0%; Score 14; DB 10; Length 203;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atcgccagatttc 16
    |||
Db 63 ATGCCAGCATTTTC 50

RESULT 39
BG589383 203 bp mRNA EST 12-APR-2001
LOCUS     EST497225 P. infestans-challenged leaf Solanum tuberosum cDNA clone
DEFINITION BPI11H7 5' sequence, mRNA sequence.
ACCESSION  BG589383
VERSION     BG589383.1 GI:13607523
KEYWORDS   EST.
SOURCE     potato.
ORGANISM   Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
            1 (bases 1 to 203)
REFERENCE   Zhang,P., Hernandez,M., Torquvist,C.-E., Wirtz,U., Loukolanov,A.,
            Rangel,P., Haberlandtch,G.T., Cho,J., Chienlingo,A., Bougri,O., Buell
            ,C.R., Ronning,C.M., Helgeson,J. and Baker,B.
            Generation of ESTs from potato leaves Challenged with Phytophthora
            infestans, Incompatible Reaction
            Unpublished (2000)
COMMENT    Contact: Cathy Ronning
            The Institute for Genomic Research
            For clone info: please contact Research Genetics, Libraries
            Division tel 1-800-711-6195, email cda@resgen.com
            Seq primer: M13P-R
            Location/Qualifiers
            1..203
            /organism="Solanum tuberosum"

```

```

/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="BPI11H7"
/clone_lib="P. infestans-challenged leaf"
/tissue_type="leaf"
/dev_stage="6 week old"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Whole plants were challenged with 450,000
sporangia/ml P. infestans US-1(US 940501) in Biotron
(Madison, Wisconsin). Leaf tissue was collected at 1, 2,
5, 12, and 24 hours post-challenge and frozen in liquid
nitrogen immediately upon removal. Kennebec plants showed
no signs of HR. Karahdin plants (susceptible to P.
infestans US-1) were used as controls and showed
infection. NOTE: We cannot exclude the possibility that
this sequence is actually derived from Phytophthora rather
than potato."
BASE COUNT   67 a 38 c 68 g 30 t
ORIGIN
Query Match 56.0%; Score 14; DB 11; Length 203;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcgccagatttc 17
    |||
Db 93 TCGCCAGCATTTTC 80

RESULT 40
BG588891 205 bp mRNA EST 10-APR-2001
LOCUS     RH1Z2_57_H09.b1.A003 Rhizome2 (RH1Z2) Sorghum proproquum cDNA, mRNA
DEFINITION RH1Z2_57_H09.b1.A003 Rhizome2 (RH1Z2) Sorghum proproquum cDNA, mRNA
sequence.
ACCESSION  BG588891
VERSION     BG588891.1 GI:13587889
KEYWORDS   EST.
SOURCE     Sorghum proproquum.
ORGANISM   Sorghum proproquum.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
            1 (bases 1 to 205)
REFERENCE   Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
            ,L.H.
            An EST database from Sorghum: Sorghum proproquum rhizomes
            Unpublished (2000)
COMMENT    Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Seq primer: JEN REV
            High quality sequence start: 4
            High quality sequence stop: 167
            PolyA-No.
            Location/Qualifiers
            1..205
            /organism="Sorghum proproquum"
            /db_xref="taxon:132711"
            /clone_lib="Rhizome2 (RH1Z2)"
            /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
            Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
            from poly-A RNA in the cloning vector Lambda Zap II.
            Clones to be sequenced were prepared by mass excision."
BASE COUNT   60 a 45 c 52 g 48 t
ORIGIN

```

Query Match 56.0%; Score 14; DB 11; Length 205;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 3 atccgcagcatttc 16
 |||||
 Db 141 ATCCGCCAGCATTTC 154

RESULT 41
 A0931496 211 bp DNA GSS 21-DEC-1999
 LOCUS RPCI-23-266P9.1J RPCI-23 Mus musculus genomic clone RPCI-23-266P9,
 DEFINITION DNA sequence.
 A0931496
 ACCESSION A0931496.1 GI:6620510
 VERSION
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 211)
 Zhao, S., Nieman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinel,
 B., Levins, M., Megann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.,
 and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 CONTACT: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pleter de Jong
 (pleter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/order/frame.htm)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
 http://www.tigr.org/tdb/bac/ends/mouse/bac_end_intro.html
 Plate: 266 row: F column: 9
 Seg primer: SP6
 Class: BAC ends.
 FEATURES
 Location/Qualifiers
 1..211
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-266P9"
 /clone_1lb="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:
 EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methylase. Size
 selected DNA was cloned into the pBAC3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."
 BASE COUNT 82 a 38 c 28 g 62 t 1 others
 ORIGIN

Query Match 56.0%; Score 14; DB 13; Length 211;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 gccacatttcatt 19
 |||||
 Db 159 GCCACATTTCGAT 172

RESULT 42

A1329089/c 231 bp mRNA EST 28-DEC-1998
 LOCUS a9q12ne.r1 Neurospora crassa evening cDNA library Neurospora crassa
 DEFINITION cDNA clone a9q12ne 3', mRNA sequence.
 A1329089
 ACCESSION A1329089.1 GI:4065648
 VERSION
 KEYWORDS EST.
 SOURCE Neurospora crassa.
 ORGANISM Neurospora crassa
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Sordariales; Sordariaceae; Neurospora.

REFERENCE 1 (bases 1 to 231)
 Zhu, H., Lai, H., Kuper, D., Dunlap, J.C. and Roe, B.A.
 Two Neurospora crassa EST Databases
 Unpublished (1998)
 Other ESTs: a9q12ne.f1
 CONTACT: Bruce A. Roe, University of Oklahoma, broeou@ou.edu
 Department of Chemistry and Biochemistry
 Advanced Center for Genome Technology, University of Oklahoma
 620 Parrington Oval, Norman, OK 73019, USA
 Tel: 405 325 4912
 Fax: 405 325 7762
 Email: broeou@ou.edu

We anticipate the future release of the cDNA clones to the Fungal
 Genetics Stock Center
 Seq primer: Universal Reverse Primer
 High quality sequence stop: 179.
 FEATURES
 source
 Location/Qualifiers
 1..231
 /organism="Neurospora crassa"
 /strain="Strain 30-7 (bd; A)"
 /db_xref="taxon:5141"
 /clone="a9q12ne"
 /clone_1lb="Neurospora crassa evening cDNA library"
 /tissue_type="tissue harvested following 22hr growth in
 dark"

/note="Vector: pBluescript SK-; Site_1: XbaI; Site_2:
 EcoRI; See: Bell-Pedersen, D., et al. PNAS 93:11096,1996.
 5' end of cDNA cloned into XbaI site of pBluescript; 3'
 end of cDNA cloned into EcoRI site of pBluescript."
 BASE COUNT 50 a 71 c 39 g 71 t
 ORIGIN

Query Match 56.0%; Score 14; DB 10; Length 231;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gccacgcagcatt 14
 |||||
 Db 57 GCATCCGCCAGCATT 44

RESULT 43
 A1902416 242 bp mRNA EST 30-MAR-2000
 LOCUS A1902416/c
 DEFINITION CM-BR006-050299-125 BR006 Homo sapiens cDNA, mRNA sequence.
 A1902416
 ACCESSION A1902416.1 GI:6492803
 VERSION
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 242)
 Dias, Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags

.....

Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:
Rosidae: eurosids I: Fabales; Fabaceae; Papilionoideae; Phaseoleae;

REFERENCE

1 (bases 1 to 271)

AUTHORS

Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna,
A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers,
Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk,
R., Ritzer, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann,
R., Waterston, R. and Wilsson, R.

TITLE

Public Soybean EST Project

JOURNAL

Unpublished (1999)

COMMENT

Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: estewatson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com

Trace considered overall poor quality
Seq primer: -40RP from Gibco

High quality sequence stop: 1.

Location/Qualifiers

FEATURES

1..271

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-c1017-5291"

/clone_1lb="Gm-c1017"

/tissue_type="vegetable buds of field grown plants"

/lab_host="X110-Gold"

/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:

XhoI; This cDNA library was constructed from mRNA isolated

from vegetable buds of field grown plants. The cDNA

library was prepared using the Stratagene pBluescript II

XR library construction kit. Complementary DNA was

synthesized from mRNA using a primer consisting of a poly

(dT) sequence with a XhoI restriction site. EcoRI adapters

were ligated to the blunt-ended cDNA fragments followed by

XhoI digestion. The cDNA fragments were directionally

cloned into the EcoRI-XhoI restriction site of the

pBluescript vector. The ligated cDNA fragments were

transformed into X110-Gold host cells. This library was

constructed by Dr. Randy Shoemaker and Dr. John

Erpelting."

BASE COUNT 88 a 52 c 67 g 64 t

ORIGIN

Query Match 56.0%; Score 14; DB 10; Length 271;

Best Local Similarity 100.0%; Pred No. 2.7e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ccagcattcgatc 20

|||||

Db 264 CCAGCATTCGATC 251

Search completed: December 26, 2001, 14:10:27
Job time: 7074 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:52:48 ; Search time 1757.07 Seconds
(without alignments)
234.726 Million cell updates/sec

Title: US-09-396-196f-8
Perfect score: 25
Sequence: 1 gcatttcgacccctgcgcagtcag 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 9

Total number of hits satisfying chosen parameters: 193545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: gb.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.om:*
20: em.or:*
21: em.ov:*
22: em.pat:*
23: em.ph:*
24: em.pl:*
25: em.ro:*
26: em.sts:*
27: em.sy:*
28: em.un:*
29: em.vi:*
30: em.htgo.hum:*
31: em.htgo.hum:*
32: em.htgo.hum:*
33: em.htg.inu:*
34: em.htg.inu:*
35: em.htg.rod:*
36: em.htg.other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	6	AR029499 Sequence
2	25	100.0	1041	6	AR034916 Sequence
3	25	100.0	1084	6	AR034916 Sequence
4	25	100.0	1121	6	E00893 Genomic DNA
5	25	100.0	5793	6	E00893 Genomic DNA
6	25	100.0	5872	6	A38246 Sequence 1
7	25	100.0	5872	6	A38251 Sequence 1
8	25	100.0	5872	6	A38251 Sequence 1
9	25	100.0	5872	6	A38251 Sequence 1
10	25	100.0	5872	6	A38251 Sequence 1
11	25	100.0	5872	6	A38251 Sequence 1
12	25	100.0	11022	1	AE000180 Sequence
13	25	100.0	13501	1	AE005258 Sequence
14	25	100.0	297816	1	AF002553 Sequence
15	21	84.0	128	1	ECOB10B
16	21	84.0	5526	1	AF250776 Sequence
17	16	64.0	2980	1	MTC1429B
18	16	64.0	14756	1	AE006972 Sequence
19	16	64.0	42923	1	MSG813GS
20	16	64.0	43839	1	MICL458
21	16	64.0	118100	2	AC0017911
22	16	64.0	171705	3	AC007476 Sequence
23	16	64.0	266133	3	AE003822 Sequence
24	16	64.0	332450	1	MLEPRN5
25	15	60.0	728	8	TCU57884
26	15	60.0	1801	8	RICCTALD
27	15	60.0	4886	9	AB037816
28	15	60.0	6372	6	AX173175 Sequence
29	15	60.0	12542	1	AE004708 Sequence
30	15	60.0	59999	2	AL496627 Human DNA
31	15	60.0	168972	2	AP003452 Oryza sat
32	15	60.0	193530	2	AP003455 Oryza sat
33	15	60.0	66	6	AR118251 Sequence
34	15	60.0	211	6	AX174645 Sequence
35	14	56.0	483	8	AF140181 Filobasid
36	14	56.0	509	8	AF140179 Filobasid
37	14	56.0	513	8	AF140176 Filobasid
38	14	56.0	526	8	AF140173 Filobasid
39	14	56.0	528	8	AF140183 Filobasid
40	14	56.0	529	8	AF140164 Filobasid
41	14	56.0	530	8	AF140165 Filobasid
42	14	56.0	530	8	AF140169 Filobasid
43	14	56.0	530	8	AF140172 Filobasid
44	14	56.0	530	8	AF140175 Filobasid
45	14	56.0	531	8	AF140160 Filobasid
46	14	56.0	532	8	AF140158 Filobasid
47	14	56.0	532	8	AF140159 Filobasid
48	14	56.0	532	8	AF140161 Filobasid
49	14	56.0	532	8	AF140162 Filobasid
50	14	56.0	532	8	AF140166 Filobasid
51	14	56.0	532	8	AF140167 Filobasid
52	14	56.0	532	8	AF140168 Filobasid
53	14	56.0	532	8	AF140170 Filobasid
54	14	56.0	532	8	AF140171 Filobasid
55	14	56.0	532	8	AF140174 Filobasid
56	14	56.0	534	8	AF140163 Filobasid
57	14	56.0	534	8	AF140177 Filobasid
58	14	56.0	1302	6	AX121194 Sequence
59	14	56.0	1425	6	AX063861 Sequence
60	14	56.0	1556	3	AX063861 Sequence
61	14	56.0	1618	1	AF327063 Corynebacter
62	14	56.0	2000	8	AF174649 Colletotrich
63	14	56.0	2014	6	AX174662 Sequence
64	14	56.0	2104	6	AR016483 Sequence
65	14	56.0	2104	6	E12594 Sequence
66	14	56.0	2104	22	E11273 DNA encodin
67	14	56.0	2194	3	AF004667 Trypanosoma
68	14	56.0	2866	8	AX113254 Sequence
69	14	56.0	3310	8	AF096871 Zea mays
70	14	56.0	3810	1	PA091631 Pseudomonas

```

c 71 14 56.0 4026 6 AX174571 Sequence
c 72 14 56.0 4027 6 AX174664 Sequence
c 73 14 56.0 4577 9 AK024436 Homo sapi
c 74 14 56.0 4639 8 ASGLAV36
c 75 14 56.0 7215 6 AX174569
c 76 14 56.0 7342 10 AF115319
c 77 14 56.0 9228 1 AE008125
c 78 14 56.0 10419 1 AE008107
c 79 14 56.0 10471 1 AE004722
c 80 14 56.0 11116 1 AE007292
c 81 14 56.0 11723 1 AE006454
c 82 14 56.0 12862 1 AE001634
c 83 14 56.0 14601 8 AE275632
c 84 14 56.0 18780 2 AC019734
c 85 14 56.0 25371 2 AC006728
c 86 14 56.0 29571 3 AE077541
c 87 14 56.0 30873 1 AE002186
c 88 14 56.0 31557 1 AC004267
c 89 14 56.0 33328 2 AC019551
c 90 14 56.0 40386 2 AC087419
c 91 14 56.0 42854 8 SPAC1745
c 92 14 56.0 61566 8 AL139346
c 93 14 56.0 63282 8 NC68B2
c 94 14 56.0 64786 2 AC017805
c 95 14 56.0 66686 6 AX174654
c 96 14 56.0 77298 8 AB011483
c 97 14 56.0 80646 2 AC013962
c 98 14 56.0 85221 2 AC019808
c 99 14 56.0 89009 1 BS275208
c 100 14 56.0 92958 2 DMBR11017

```

ALIGNMENTS

```

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE
ORGANISM
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 1041)
TITLE Patton,D.Andrew.
JOURNAL Enhanced biotin biosynthesis in plant tissue
FEATURES
SOURCE Location/Qualifiers
1..1041
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 1.7e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 gcatlccgactcctgcagtgacg 25
Db 99 GCATTTCGATCCTCGTCAGTGCAG 123

RESULT 2
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.

```

```

ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 1041)
TITLE Patton,D.A.
JOURNAL Transgenic plants having increased biotin content
FEATURES
SOURCE Location/Qualifiers
1..1041
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 1.7e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 gcatlccgactcctgcagtgacg 25
Db 99 GCATTTCGATCCTCGTCAGTGCAG 123

```

```

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION BiOB gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.

```

```

REFERENCE 1 (bases 1 to 1084)
AUTHORS
JOURNAL
FEATURES
SOURCE Patent: GB 2216530-A 16 11-OCT-1989;
Location/Qualifiers
1..1084
/gene="BiOB"
/gene="BiOB"
/gene="BiOB"
/gene="BiOB"
/codon_start=1
/transl_table=1
/protein_id="CAA00965.1"
/db_xref="GI:490219"
/transl_table="MAHRPMTLSOYVLEFEPKPLDLLPEAOVYHROHDPPOVYST
LASTKGCAPEDCKYCPROSSRYRTGLFAEELMVEVOYLESARAKAAGSTRMGAM
KNHFERMPYLEOVOCVAKMGLBACHTLTLESQARLANAGLDYVNHNDLTPSEF
VGNLITRTYOEKLDLEKVDAGIKVCSGIVIGELVYKDRAGLLQLANLPTPEP
VPINMLKVKGTPLADNDVDADFDIKTIVARIMPTSVYRSASRQDNEDQOAMC
FMAGANSIFYGKILITTPNEDEKDLQLEFRKLGILNPQAVLAGDNEQOORLEQALMT
PDTDEYYNAAL"

```

```

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 1084;
Best Local Similarity 100.0%; Pred. No. 1.7e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 gcatlccgactcctgcagtgacg 25
Db 122 GCATTTCGATCCTCGTCAGTGCAG 146

```

```

RESULT 4
LOCUS E00893 1121 bp DNA
DEFINITION Genomic DNA encoding biotin Synthetase.

```

ACCESSION E00893
 VERSION E00893.1 GI:2169154
 KEYWORDS JP 1986149091-A/1.
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
 REFERENCE 1 (bases 1 to 1121)
 AUTHORS Hiroo, Y., Kojima, T. and Kimura, H.
 TITLE DUPLICATION OF BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND PRODUCTION OF BIOTIN
 JOURNAL Patent: JP 1986149091-A 1 07-JUL-1986;
 COMMENT NIPPON SODA CO. LTD
 OS Escherichia coli
 PN JP 1986149091-A/1
 PD 07-JUL-1986
 PF 24-DEC-1984 JP 1984272605
 PI HIROO, YOSHITAKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
 C12N15/00, C12N1/20, C12P13/18, (C12N1/20, C12R1:19), (C12P13/18, PC C12R1:19);
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC *Source: strain=Escherichia coli Nsl01;
 CC feature is identified by experimental;
 FH key Location/Qualifiers
 FT CDS 42..1079
 /product="biotin synthetase".
 FEATURES
 source Location/Qualifiers
 1..1121
 /organism="Escherichia coli"
 /db_xref="taxon:562"
 BASE COUNT 289 a 296 c 325 g 211 t
 ORIGIN
 Query Match 100.0%; Score 25; DB 6; length 1121;
 Best Local Similarity 100.0%; Pred. No. 1,7e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 gcattcgatcctcgtagtgacg 25
 ||||||||||||||||||||
 Db 140 GCATTGCACTCTCGTCAAGTGCAG 164
 RESULT 5
 ECOBIO 5793 bp DNA EC⁺ 28-FEB-1994
 LOCUS E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
 DEFINITION (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioc
 protein, and delthiobiotin synthetase (bioD), complete cds.
 ACCESSION J04423
 VERSION J04423.1 GI:145422
 KEYWORDS 7,8-diamino-pelargonic acid aminotransferase;
 7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
 bioC gene; bioD gene; bioF gene; biotin synthetase; delthiobiotin
 synthetase;
 SOURCE Escherichia coli (strain K-12) DNA.
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
 REFERENCE 1 (bases 1 to 5793)
 AUTHORS Otsuka, A.D., Buonocristiani, M.R., Howard, P.K., Flamm, J. and Johnson, O.
 TITLE The Escherichia coli biotin biosynthetic enzyme sequences
 JOURNAL J. Biol. Chem. 263, 19577-19585 (1988)
 MEDLINE 89066784
 COMMENT Draft entry and computer-readable sequence [1] kindly submitted by A.Otsuka, 09-NOV-1988.
 FEATURES Location/Qualifiers

source 1..5793
 /organism="Escherichia coli"
 /strain="K-12"
 /db_xref="taxon:562"
 complement(98..574)
 CDS
 /note="ORF 1"
 /codon_start=1
 /transl_table=11
 /protein_id="AA23513.1"
 /db_xref="GI:45168"
 /translation="MKLISNDLRGDKLPKRRHVENMGVYDGDNISPHLADVPAGTK
 SEVYVVCYDPDAPPTGSGMHWVNLPAQTRVLPQFGSLVAMPDGVLOTPTDCKTG
 YDGAAPKGETHRYLEFTHALDIERIDVDEGASAMVGFNVHFLSLASITAFMS"
 complement(633..1925)
 /gene="bioA"
 complement(633..1925)
 /codon_start=1
 /gene="bioA"
 /transl_table=11
 /protein_id="AA23514.1"
 /db_xref="GI:457106"
 /translation="MTTDLAFDQRIHPTTSMTPSPVYPVSAECCCELLISGR
 LIVGMSWMAALHGYNHQPLNAAKMSQIDAMSHVFGITPAATELCRLVAMGRN
 ALECVFLADSGVAVEVAMKMALOYWAKGEARQPLFRNGYHGDITGAVISVCPDN
 SMHSLKMGVLPENLEAPAPQSRMDGEMDRDVGAPRLMAARHEIAVILEPIVGA
 GCMRWYHPMLKRIKICDREGILLIADIEAFGRITGKLPACHEAETAPILLEGKA
 LTGGTMTLSATLTTRVAVETISNGRAGCTMHPGMPACAAANASATILLEGSDMO
 QQVADIEVLRQDLAPARDAEVAADVIVGALGVVETTHPVMAALQKFPVEQGWIR
 PFGKLIYMPYIILPQQLQRLTAANRAVODETEFCQ"
 2012..3052
 /gene="bioB"
 2012..3052
 /gene="bioB"
 /codon_start=1
 /transl_table=11
 /product="biotin synthetase"
 /protein_id="AA23515.1"
 /db_xref="GI:145425"
 /translation="MAHRPMTLSQVTELEFEKPLDLLFEAQOVHROHPDPOVST
 LLSIKTGACPEDCKYCPQTSRKTKGLEARLHEVEYDESARAKAAGSTRCKMAAM
 KNPHERDMPLLEQWGVKAKMGLEACMTLGTLSQSAORLANAGDYIYHNIDTSPER
 YGNIITTRVYQERLDLTLEKVRDAGIKVCGGIVGJGVKAGLIDLANLPYTES
 VPINMLVKVGYPLADNDVDVADFRTIYAVIMPTSYVRLSAGREQMNSQTAAC
 FMAGANSIFVCKLTTPNPBEDKDLQLPFKLGLINQCAVAVLAGDNEQOQRLQALMT
 PPRDEVYNAAL"
 3049..4203
 /gene="bioF"
 3049..4203
 /gene="bioF"
 /codon_start=1
 /transl_table=11
 /product="7-keto-8-amino pelargonic acid synthetase"
 /protein_id="AA23516.1"
 /db_xref="GI:145426"
 /translation="MSMOEKINAALDARRAADALRRRYPVAGAGRMVADDDQYLINE
 LSNDDYGLSHHPOIIRAMOGAEOFGISGSGSVAGSYVHQALEELAEMLGYSRA
 LFTISGFAANDAVIAMAAMEKRIADRSASLILPAASLSPQALRRFANDYTHIR
 LLASPCGQOMVVTVECFVSMDDGSAPLAETIOQVTOOHNMIMLWDDHAFVGEORG
 SCHLQKVRPELVAVTFEGKGVGCAVLCSSIVADILQFARLLIYSTMPAPQANAL
 RASLAVLRDEGDARREKLAALLTRFRAGVDLPFTLADSCSAIOPLVIGDSRALQL
 AEKLRQGCWVTAIRPTVPAGTARLLTLTAHEMODIDRLLEVLHGNG"
 4190..4945
 /gene="bioC"
 4190..4945
 /gene="bioC"
 /codon_start=1
 /transl_table=11
 /protein_id="AA23517.1"
 /db_xref="GI:145427"
 /translation="MATVKNQALIAAAGRAAHYEDHADLOROSADALLAPQRYT
 HVIDAGCGPCPMKSRHMERRAOYVATLDSPPMLVQARODADHYLAGDIESLPLATA
 TFDLAWSNLAVQWCGNLSIALRELRYVPRKGVVATPTLVQSSLPERHAKMAQVADIERP

```

HANSRLPDETEOSLNGVYOHHTIOPITLWEDDALSMRSKLGATGHLHEGRDPRIL
TRSOLORLQALAMPQOQGRPYTLTHLFLGVARE
4938..5597
/gene="bioc"
/db_xref="GI:2294847"
4938..5597
/gene="bioc"
/codon_start=1
/transl_table=11
/product="gethobiotin synthetase"
/protein_id="CAA23518.1"
/db_xref="GI:145428"
/translation="MSKRYFVTGDEEVGKTVASCALLQAKKAGYRTAOYKPVASCS
EKTPPEGLNSDALQPNSSLOLDYATVNPYTAEPSPHIIAOGPRPESLVMASG
LRALEQADWLVAGAGWPTPLSDPTTFADWVTOPLPYIIIVGVKLCGGINIAMJTA
OVIQHAGLTLGAWANDVTPPKRIHAYMTTLTRMTFPRCWERSPOLQIKHROPES
T"
BASE COUNT      1363 a 1554 c 1631 g 1245 t
ORIGIN           4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match      100.0%; Score 25; DB 1; Length 5793;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatcgcacccgcgcagcagcag 25
|||||
Db 2110 GCATTTCGATCCCTCGTCAGTCAG 2134

RESULT 6
LOCUS           A38246          5872 bp      DNA
DEFINITION      Sequence 1 from Patent WO9408023.
ACCESSION       A38246
VERSION         A38246.1 GI:2294844
KEYWORDS
SOURCE          Escherichia coli.
ORGANISM        Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
                Escherichia.
REFERENCE       1 (bases 1 to 5872)
AUTHORS        Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLES         BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL         Patent: WO 9408023-A 1 14-APR-1994;
                LONZA AG (CH)
COMMENT         Other publication PL 308301 950724
                Other publication CA 2145400 940414
                Other publication AU 4820293 940426
                Other publication HU 71781 960228
                Other publication SK 42095 951108
                Other publication CZ 9500809 950913
                Other publication FI 951547 950331
                Other publication JP 8501694T 960227.
                Location/Qualifiers
                1..5872
                /organism="Escherichia coli"
                /strain="DSM498"
                /db_xref="taxon:562"
                /clone="PB030A-15/9"
                1..96
                /function="PROMOTER PTAC"
                /evidence=experimental
                23..28
                /standard_name="PROMOTER PTAC"
                /evidence=experimental
                45..49
                /standard_name="BIOB RBS NO.9"
                /evidence=experimental
                117..1157
                /gene="BIOB"
                117..1157
                /gene="BIOB"
                /gene="BIOB"

FEATURES
source
promoter
-35-signal
-10-signal
RBS
gene
CDS

```

```

stem_loop      5583..5605
terminator     5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT    1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gcattcgatccctgcagtgcaq 25
|||||
Db 215 GCATTTCGATCCTGTCAGTGCCAG 239

RESULT 7
LOCUS      A38251      5872 bp      DNA
DEFINITION Sequence 6 from Patent WO9408023.
ACCESSION  A38251
VERSION     A38251.1 GI:2294849
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL     Patent: WO 9408023-A 6 14-APR-1994;
LONZA AG (CH)
OTHER PUBLICATION PL 308301 950724
OTHER PUBLICATION CA 2145400 940414
OTHER PUBLICATION AU 4820293 940426
OTHER PUBLICATION HU 71781 960228
OTHER PUBLICATION SK 42095 951108
OTHER PUBLICATION CZ 9500809 950913
OTHER PUBLICATION FI 951547 950331
OTHER PUBLICATION JP 8501694T 960227.
Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/organism="BIOF"
1154..2308
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAA02329.1"
/db_xref="GI:2294850"
/translation="MSWQEKINALARAPADALRRPYAAGAGRMILVADDDVOYLF
SSNDYDLSHHPOLIIRAMDOGAEQFCIGSGSCHVSCYSYHDALEELAEMLGYSKA
LLFTSGFANQAVYIAAMMAKEDRIARURLSHASLLEASLSPSOLRFANVDVTHLAR
LSPGPGQOMVYVEGVFSMKGDSAPLAEIQVTOQINGWLMDADAGTGVTIGEOQGR
SCWLOKVKPELLVTEGKFGVSGAAVLCSSTVADYLLQFARHLIYSTMPPAQOAL
RASLAVIRSDGEGDAREKELALITRFRAGVODLPTLIADSCSAIOLPLTVGDSRAIOL
AEKROGCGWTAIRPPPTVPAQTAARLRLTLTAHHEMDIDRLLEVLHGNC"
3030..3045
/standard_name="BIOB RBS"
3043..3753
/gene="BIOB"

```

```

CDS
3043..3753
/gene="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="DPR SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYEVTGDTVEGKTVASCALLQAKAGYRTAGYKPYVSGS
EKTEGELNSDALALORNSSLQDYATVNPYFAEPSPHISAQESRPTESLYMSAG
LRALQODWVTVLGAGCGMPTPLSDTFEADWVVOEOLPVLYGVYGLGGINHMLTA
OVIOHAGITLACGWVANDYTPPGKRHAEMTTLTPMIRAPILGELPMLAENPENAAATGK
YINLAFVASTLTGFTSR."
YINLAFVASTLTGFTSR.

BASE COUNT    1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gcattcgatccctgcagtgcaq 25
|||||
Db 215 GCATTTCGATCCTGTCAGTGCCAG 239

RESULT 8
LOCUS      A93674      5872 bp      DNA
DEFINITION Sequence 1 from Patent EP0798384.
ACCESSION  A93674
VERSION     A93674.1 GI:6741862
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch,O. and Brass,J.
TITLE      Biotechnological method of producing biotin
JOURNAL     Patent: EP 0798384-A 1 01-OCT-1997;
LONZA AG (CH)
Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A-15/9"
1..96
/function="PROMOTER PTAC"
/evidence=experimental
23..28
/standard_name="PROMOTER PTAC"
/evidence=experimental
45..49
/standard_name="BIOB RBS NO.9"
105..119
/standard_name="BIOB"
117..1157
/evidence=experimental
117..1157
/gene="BIOB"
117..1157
/organism="BIOB"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRRPWTLSQVTELFEXPLLDLLEAOQVHRQHFDPROYVST
LTIKTAGCPEDCKYCCPOSSRYKTGLEAFRLMEYVQVLESARAKAAGSTRFCMAAM

```

```

KNHERDMPYLEOMVOCVAMGL EACHTLCTLS ESOAORLANAGLDTYNNHNDTSPR
YGNITTRTYQERLDITLEKVDGKICSGGIVGLBETVADRAGLLJOLANETPTPEP
VPINMLVKVGTPLADNDVDADFDIETKTAVARIIMETSYVRLSAREQONEDTOAMC
FMAGANSIFYGCKLITTPNEPEDKDLQLEFRKGLINPOQAVLAGDNEOQORLEQALMT
PDTDEYVNAAL"
2284. .2297
RBS
/standard_name="BIOC RBS"
2295. .3050
gene
/gene="BIOC"
2295. .3050
CDS
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVKNQALAAFGRAAHYEQHADLOROSADALLAMLPOKVI
HVLDAQCGPGMSRHRERHAYTALDISPMI.VOAROKDAADHYACDIESIPLATA
TFPLAMSNLAVMCGNLTALRELYRVPRKGVYATTLVOSLPELHOMQAVDERP
HANRFLPDELEOSLNGVYQHIIQPTTMEFDALSAKSLKIGATHLHEGRDPR11.
TRSQLORJOLAMPQOQGRVPLTYHLEGLVARE"
3742. .3752
RBS
/standard_name="BIOA RBS"
3750. .5039
gene
/gene="BIOA"
3750. .5039
CDS
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONANONATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLAFDORHIMHPYTSMTSP.PVYVYSAEGCELLISDCK
LNDGSSMAATIGYNNHLPOLNAAKSOITDAMSVMPGTTTHADATLCKRLVATPDP
LECYFLADSSVAVEYAMKVALQYWOAKGEARQRLTFRNGYHGTGFGMSVCDPDS
MBSLWKGYLEPNI.FAPAPQSRMDGEMDERDMGFARLMAAHRHEIAVITIEYQAG
GMSRWYEWLKRIRKICDREGILLADEIATGFRGKLFACENAEIADILLGKAL
TGCTMTLSATLITREVAETIISNGEACGPMHGTGMPNPLACAAANASLLLSGDMOQ
OVADIEVLOREOLAPARDAEMVADVRLGAGVETTHPYVMAALQKPFVEGVMIRP
FKGLIIMPYILLPQOLRLTAANRAVODEFFEQ"
5088. .5100
RBS
/standard_name="ORF1 RBS"
5098. .5574
gene
/gene="ORF1"
5098. .5574
CDS
/gene="ORF1"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDGDKLPHRIVENGMGYDGNISPHLAMDVPAGTK
SPVYTCYDDAPATGSGMHWVYVNLPAIDRVLPQFGSGLVAMPDQIDITRDPRKTG
YDGAAPPKGETHTHTYITVHALDIERIDVDEGASGAVGVNHFHSLASITAMS"
5583. .5605
stem_loop
5583. .5644
terminator
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

```

Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 1.8e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 gcaattcgatcctcgatcgatcgag 25
Db 215 GCATTCGATCCGCTGACGTGCAG 239

RESULT 9
A93679
LOCUS A93679 5872 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 6 from Patent EP0798384.
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE
ORGANISM Escherichia coli.
Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O. and Brass,J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
source location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONANONATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSMOEKINNALDARRADALRRRYPAVAGRWLVADDDROYLNF
SSNYVIGSHHPOLIRAMQOAGEGIGSGSGVGVHOLAEELAEWIGYSRA
LLTISCFRANQAVIAAMAKDRITAIRLISASILLEASISPOLRFRANDVTHLAR
LLASPCRGQMWVTEGVFSMDGSAPIAEIOVYQHNGLMWDADGTGVTGEGGSG
SCWLOKVPBELLVVTFKGFQVSGAAVCSYVADYLLQFARHLIYSTSPPOAOAL
RASLAVIRSDGDAIREKLAALITRFAGVODLPFTLADSCSAIOPLVGDNSRAIOL
AEKLRQOGCWYTAIRPTVPAGTARLRLTLAAHEMODIDRLLEVLHGNG"
3030..3045
RBS
/standard_name="BIOD RBS"
3043..3753
gene
/gene="BIOD"
3043..3753
CDS
/gene="BIOD"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYVYTGDPFVEGKTVASCAILQAKAKAGRTAGYKRVASG
EKPRPEGLRNSDALQINSSLOLDIATVYNTYTAERTSPHITISAOGRPTESLYMSAG
LRALQADADVIVGAGGMFTPLUSDITTEADWYQOLPVLIVGVKLGCTINAMULTA
QVIOHAGLITLAGVAVANDVTPPGKKRHAETMTLIRMLPAPLGEIPMLAENPENAAATG
YINLAFVDASTLGFPSRL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

```

Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 1.8e-05;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcaattcgatcctgcagtgacg 25
 |||||||||||||||||||
 Db 215 GCATTCGATCCTGCTGACGTGCAG 239

RESULT 10
 ARI01809 5872 bp DNA PAT 14-FEB-2001
 LOCUS Sequence 1 from patent US 6083712.
 DEFINITION ARI01809
 ACCESION ARI01809.1 GI:12812607
 VERSION
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
 FEATURES Location/Qualifiers
 source 1..5872
 /organism="unknown"

BASE COUNT 1318 a 1552 c 1695 g 1307 t

Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 1.8e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcaattcgatcctgcagtgacg 25
 |||||||||||||||||||
 Db 215 GCATTCGATCCTGCTGACGTGCAG 239

RESULT 11
 ARI01810 5872 bp DNA PAT 14-FEB-2001
 LOCUS Sequence 6 from patent US 6083712.
 DEFINITION ARI01810
 ACCESION ARI01810.1 GI:12812608
 VERSION
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
 FEATURES Location/Qualifiers
 source 1..5872
 /organism="unknown"

BASE COUNT 1318 a 1552 c 1695 g 1307 t

Query Match 100.0%; Score 25; DB 6; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 1.8e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcaattcgatcctgcagtgacg 25
 |||||||||||||||||||
 Db 215 GCATTCGATCCTGCTGACGTGCAG 239

RESULT 12
 AE000180 11022 bp DNA BCT 01-DEC-2000
 LOCUS Escherichia coli K12 MG1655 section 70 of 400 of the complete
 DEFINITION genome.
 ACCESSION AE000180 U00096

VERSION AE000180.1 GI:1786988
 KEYWORDS
 SOURCE Escherichia coli K12.
 ORGANISM Escherichia coli K12
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1 (bases 1 to 11022)
 AUTHORS Blattner, F.R., Plunkett, G., III, Bloch, C.A., Perna, N.T., Burland, V.,
 Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
 Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
 Mau, B. and Shao, Y.
 TITLE The complete genome sequence of Escherichia coli K-12
 JOURNAL Science 277 (5331), 1453-1474 (1997)
 MEDLINE 97426617
 PUBMED 9278503

REFERENCE 2 (bases 1 to 11022)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecolli@genetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459

REFERENCE 3 (bases 1 to 11022)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecolli@genetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459

REFERENCE 4 (bases 1 to 11022)
 AUTHORS Plunkett, G., III.
 TITLE Direct Submission
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 This sequence was determined by the E. coli Genome Project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHGR). The entire sequence was independently
 determined from E. coli K12 strain MG1655. Predicted open reading
 frames were determined using Genemark software, kindly supplied by
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
 have been correlated with genetic loci are being annotated with CG
 Site Nos., unique ID nos. for the genes in the E. coli Genetic
 Stock Center (CGSC) database at Yale University, kindly supplied by
 Mary Berlyn. A public version of the database is accessible
 (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and
 its annotations are periodically updated; this is version M54. No
 sequence changes. Annotation updates: updated gene identifications
 and products; all new functional assignments courtesy of Monica
 Riley; added promoters, protein binding sites, and repeated
 sequences described in reference 1. The unique numeric identifiers
 beginning with a lowercase 'b' assigned to each gene (protein- or
 RNA-encoding) are now designated as gene synonyms instead of
 labels. This should allow them to be searched for in Entrez as gene
 names.

FEATURES
 source 1..11022
 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="MG1655"
 /db_xref="taxon:83333"
 <1..130
 /note="REP (repetitive extragenic palindromic) element;
 contains 4 REP sequences"
 complement(147..1430)
 /gene="ybhC"
 /note="b0772"

repeat_region
 gene

CDS
 complement(147. .1430)
 /gene="ybhC"
 /function="orf: Not classified"
 /note="f427: 98 pct identical to fragment YBH_C ECOLI
 SW:46130 (300 aa) but contains 127 additional C-terminal
 residues"
 /codon_start=1
 /transl_table=11
 /product="putative pectinesterase"
 /protein_id="AAC73859.1"
 /db_xref="GI:1786989"
 /translation="MNTESVSRLALALAFGVTLTACSSPPDQPSDQTAFCGSSRPI
 LSAEAGNPDAQHYEASTLPGAAAMNSPTLLPAQDPFVGPACTQCVHTTTIOAIVD
 AALIKRNRKROXIAVMPGEYQGYTYVAAAGCITLTGTERPIDVKIGLSLDGMSIA
 DMKRDVPRGKIMPKFPMATWYDSCQSRSDSGVCSAVFMSQNNGLQNLQNTLNT
 LGSDVAGNHPAVALRTDGDQVQINNVAIIGRQNTFVNSGVQNRLETNRQPTLV
 NSYLEGDVIVSGCAVFPDTEFRVNSRTOEAVYFAPATLSNIYGYAANSREN
 AFGDVAGQSGSLDVANTNGQVYIRDSAINEGFNIAKPMADAVISNRPFAGNTGSVD
 DNDIQRNLNDTNYNRMEYNNRGVSKVAEAKK"
 complement(1507. .1535)
 /note="factor Sigma70: predicted +1 start at 806574"
 /complement(1582. .2058)
 /gene="ybhB"
 /note="b0773"
 /complement(1582. .2058)
 /gene="ybhB"
 /function="orf: Unknown"
 /note="f158: 99 pct identical to YBH_ECOLI SW: P12994"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC73860.1"
 /db_xref="GI:1786990"
 /translation="MKLISNDLRDGRKLPRIHFNCGVDCGINSPLAMVDYPACTK
 SFVYTCIDPAPATGSGMHWVNVLPADTRVLPQGFSGGLVAMPDGLQTRIDFGTQ
 YDGAAPRGSTHAYITFVHALDIERIDVEGASCAVGFVHFHSLASAITAMFS"
 complement(2108. .2124)
 /note="central position to predicted promoter:85.5"
 /bound_moiety="Rhas predicted site"
 complement(2117. .3406)
 /gene="bioA"
 /note="b0774"
 /complement(2117. .3406)
 /gene="bioA"
 /EC_number="2.6.1.62"
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="f429: 100 pct identical to BIOA_ECOLI SW: P12995"
 /codon_start=1
 /transl_table=11
 /product="7,8-diaminopelargonic acid synthetase"
 /protein_id="AAC73861.1"
 /db_xref="GI:1786991"
 /translation="MTTDLAFDQRHIMHPYMSPTLPVYVVSAGEGELLISGRK
 LNDGSSMAATIHGYNHNPOLNAAKMSQIDAMSHVPRGTTTAPLELCKLIVAMP
 LECVPLADSGSAVEYAKMALQYOWAKGEARORLTFENAGHGFPGAMSVCCDDNS
 MHSIMKGYLPENLEFAPQSRMDGEDEWDWGFALMAHHEIAAVITTEPVGAG
 GMRMYPEMLKRIKICDREGILLINDELTATGFRGKLFACENAEIADIICTGKAL
 TGGTMTLSATLTREVAETISNGEACGFHAGPTFMCNPLACAANAASLALIESGMQO
 OVADI EVOLEQLEAPADAENADVAVLCAIGVETTHVYVMAALOKPFVEGVWIRP
 FGLIYLPMPYIILPOOLRIITAAVRAAVDDETFQO"
 complement(2193. .2221)
 /gene="bioA"
 /note="factor Sigma70: predicted +1 start at 807260"
 /complement(2193. .2221)
 /gene="bioA"
 /note="central position to bioA promoter: -20"
 /bound_moiety="BioB documented site"
 /complement(2193. .2221)
 /note="central position to predicted promoter: -20, genetic
 evidence for the site"
 /bound_moiety="BioB predicted site"
 /complement(2193. .2221)
 /note="central position to predicted promoter:50"

protein_bind
 /bound_moiety="BioB predicted site"
 complement(3411. .3450)
 /note="central position to bioA promoter:50"
 /bound_moiety="BioB documented site"
 /complement(3411. .3450)
 /note="factor Sigma70: promoter bioB; documented +1 at
 808525"
 /complement(3447. .3473)
 /note="factor Sigma70; promoter bioA; documented +1 at
 808515"
 /complement(3493. .4533)
 /gene="bioB"
 /note="b0775"
 /complement(3493. .4533)
 /gene="bioB"
 /EC_number="2.8.1.1"
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="f0346: 99 pct identical to BIOB_ECOLI SW: P12996"
 /codon_start=1
 /transl_table=11
 /product="biotin synthesis, sulfur insertion?"
 /protein_id="AAC73862.1"
 /db_xref="GI:1786992"
 /translation="MAHRPMTLSQVTELEFEKPLDLLFEAQVYHROHFPDQYVST
 LLITKYGAPEDCKYCPQSSRYKIGLEAEERLEVEOYLEAKKAKAGSFPFGAM
 KNPHEDRMYLEQWQVYKAMGLACMTLGLTSSQORLANGLADYVNNHLDTPSP
 YGNTITRTYOEGLDTLEKVRDAGIKVSGGIVGIGETVRDAGLLDLOLPLPSPS
 VPINLKVKGPLADNDVDAPFIRTAIRIMPSYRISAGREONREOQAMC
 FMAGNSIFCYGCKLLTTPNPEEDKDLOLFRKILGINPQOTAVLADNDQORLEQALMT
 PDMPRYNAAL"
 /complement(4530. .5684)
 /gene="bioE"
 /note="b0776"
 /complement(4530. .5684)
 /gene="bioE"
 /EC_number="2.3.1.47"
 /function="enzyme: Biosynthesis of cofactors, carriers:
 Biotin"
 /note="f0384: 100 pct identical to BIOE_ECOLI SW: P12998"
 /codon_start=1
 /transl_table=11
 /product="8-amino-7-oxononanoate synthase"
 /protein_id="AAC73863.1"
 /db_xref="GI:1786993"
 /translation="MSWQEKINAALDARRAADALRRRYPYAGAGRWLVADRDQYNE
 SSNDYLGSHHPOLIIRAOQCAEQFGIGSGSGHVSQSVVHQALEBELLEWGYSA
 LLFTSGFANOVAVIAAMAKEDRIAAQRLSHASILEAASLPSQLRFAHNDVTHLAR
 LLASPCGOQWVTEGVSMDSAPLAELIOQVTOOHNGLVMDADRGCTGVIEOGRG
 SCWLQKVPPELLVVTFCGKFGVSGAAVLCSITVADYLLQFARHLIYSTSPRAQAL

Query Match
 Best Local Similarity 100.08; Score 25; DB 1; Length 11022;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3591 GCATTTCGATCCTCGTCAGTGTCAG 3615

RESULT 13
 AE005258
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE

AE005258 13501 bp DNA BCT 21-MAR-2001
 Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
 of 155
 AE005258 AE005174
 AE005258.1 GI:12513751
 Escherichia coli O157:H7 EDL933.
 Escherichia coli O157:H7 EDL933
 Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 I (bases 1 to 13501)

AUTHORS Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Postal, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grobbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamoudis, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A., and Blattner, F.R.

TITLE Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7

JOURNAL Nature 405 (6819), 529-533 (2001)

MEDLINE 21074935

POBED 11206551

REFERENCE 2 (bases 1 to 13501)

AUTHORS Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Postal, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grobbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamoudis, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A., and Blattner, F.R.

TITLE Direct Submission

JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

FEATURES

source

1. 13501
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="Enterohaemorrhagic"
/note="Island #36; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655; Cryptic prophage CP-933K; Includes one copy of the 13 bp direct repeat that flanks the prophage"

misc_feature

66..665
/gene="lomK"
/note="20981"
66..665
/gene="lomK"
/function="putative membrane; Other or unknown (Phage or Prophage Related)"
/note="Residues 1 to 199 of 199 are 72.86 pct identical to residues 1 to 199 of 199 from GenPept 118 : g117532789|gb|AAE63231.1|AF151091_2 (AF151091) lom prophage P-E1bA]"
/codon_start=1
/transl_table=1
/product="putative outer membrane protein of prophage CP-933K"
/protein_id="AAG55137.1"
/db_xref="GI:12513752"
/translation="MRKCAALISALICLAVSPVAPMASEHOSITLACGLHASTDAPG SDDINGINKRYEPTDGLITSTSYANADEBOKTHSDTRMHEDYRNKWFVSMAG PSVRNENFSAVAMAGVAYSRVSTFSGYFRTVDNKRKTHDVLTGSDDARY SNTSLAM GAGVQNPPEPSAVADVAYEXSGSGMDRTGFIYGVYK"

gene

719..2041
/gene="20982"
719..2041
/gene="20982"
/function="putative structure; Structural component (Phage or Prophage Related)"
/note="Residues 164 to 440 of 440 are 68.79 pct identical to residues 381 to 645 of 645 from GenPept 118 : g114585436|gb|AAD25464.1|AF125520.59 (AF125520) putative tail fiber protein [Bacteriophage 933w]"
/codon_start=1
/transl_table=1
/product="putative tail component of prophage CP-933K"
/protein_id="AAG55138.1"
/db_xref="GI:12513753"
/translation="MGMAVVOISGLVKGAGKPIQNGTJOLKARRNSTXVVVNTVASE NPDAGRYSMDEYGOYVTLVGGPPHAGTISVDEDSOGTILNDFLGAMTDEDDAR PALRRFDQVVEEAHRAHAEKKNAGEATSKRNAGISAKSAASANAADTSABEASE SARQAESAASAKKSEASASSASEAQAASLSLSDAIDAEISKRTAESAAGNARDA TTSTKARSAESAQSAOSRIADAVNRIPTVVCPGPGKGPAGPGPGKGRCE

CDS

gene

2489..3469
/gene="20985"
2489..3469
/gene="20985"
/function="orf; Other or unknown (Phage or Prophage Related)"
/note="Residues 13 to 326 of 326 are 55.73 pct identical to residues 18 to 331 of 336 from GenPept 118 : g116960367|gb|AAE33527.1 (AF170176) hypothetical protein predicted by Glimmer [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55140.1"
/db_xref="GI:12513755"
/translation="MLSPTRTFPHNSVINVOSSPCQTVSPAGKEYELKVIDEKTPILE KMEPEPNERYKDEEPIVNTKQHPYLDNTNARIISDMIGIFVGDSEVNOQTAFS KLREDFENWMIYREDVDPSMDYKRLSDIYHDICORLRTEDKRDYELNLLEKLR EISKADSLISWAKKRNHAWPDPFNALIKAGELFRCYNTKRNHGISFEGECITYD MDMLTGLKLTITAPDGLSMHVDNRDSDNTEISALIVNRSNHPALBELSPHNSKVD AHPYDGLGKGYENFPLNHNFCDEIFERNHIIIMTSQYTCSS"

CDS

gene

3530..4522
/gene="20986"
3530..4522
/gene="20986"
/function="orf; Other or unknown (Phage or Prophage Related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55141.1"
/db_xref="GI:12513756"
/translation="MKTPSLQSNFNSAPAGYCAPLAPNAENAADVLDIGKRIPL SAAIDSNVYESVIRAVDSRSRLIDQHTVDMIGNTVLDALSQTFRDVAVSYGIIHEK VHGICITRYNDEYELNEESSVKIDIDISLTCTNELYEDGCEPIFPICEAGENEDEPY VSGVAPDPSYELNEMSGELTHEIILHHVGGSSDPSGDSNITELGPIELIARVYOELG WSYVDKCYAPDEPEEHLRLRNALROAAMRHEDENRAPFERLGTISRYEASPDFT EYSVSNIGYGFIOQHDPEGLAINDLNQANOIYHGARYITTFGDVYKHNOQ"

CDS

gene

5350..6231
/gene="20989"
5350..6231
/gene="20989"
/function="orf; Other or unknown (Phage or Prophage Related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55142.1"
/db_xref="GI:12513757"
/translation="MLSPYSVNLGCSNLSLTRLTSPDNRVYLSVDAVHSDNGAOV

```

      gene
      CDS
      6462..7160
      /gene="20990"
      /note="20990"
      /function="orf: Other or unknown (Phage or Prophage
      Related)"
      /note="No significant matches"
      /codon_start=1
      /transl_table=11
      /product="unknown protein encoded by prophage CP-933K"
      /protein_id="A05143.1"
      /db_xref="GI:12513758"
      /translation="MRPTSLNLYLHQSRRSSMDIDIESLAKSSQWIKNPOLRF
      OGIDHNIYQOIEALDKIGSETGRVLNIESISRKSETVIHINSSRLGMARD
      IDENHRTGSDPFCNNAVEYPCGEGISYVDPHATVPHLLHVNHLNGERLAKES
      SRAASKYSPILLEEARTVGLGAFSEVLESENKFHEEIGMPRTSTPXSALIHDDNT
      VSUGQOVRLLPPL"
      /complement(7668..8144)
      /gene="ybhB"
      /note="20992"
      /complement(7668..8144)
      /gene="ybhB"
      /function="orf: Unknown function"
      /note="Residues 1 to 158 of 158 are 99.36 pct identical to
      residues 1 to 158 of 158 from Escherichia coli K-12 strain
      MG1655: B0773"
      /codon_start=1
      /transl_table=11
      /product="orf, hypothetical protein"
      /protein_id="A05144.1"
      /db_xref="GI:12513759"
      /translation="MKLISNDLRKDGKPLPRHAFNGMGYDGNISPHLAMDVPACTK
      SFVVTGCDPAPATGSGMHWVNVNLPADPTVLPGGSGSLVAMPDGLDTRDFEKG
      YDGAAPKGETHETIFVHALDVERIDVDGASGAVGFWVHSHLSASTAMFS"
      /complement(8203..9492)
      /gene="bhaA"
      /note="20993"
      /complement(8203..9492)

      CDS
      Query Match 100.0%; Score 25; DB 1; Length 13501;
      Best Local Similarity 100.0%; Pred. No. 1.8e-05;
      Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

      Oy 1 gcaattcgatctcgtcagatgcag 25
      |||||||||||||||||||||||||||
      Db 9677 GCATTTCGATCTCTCGTCAGTCAG 9701

      RESULT 14
      AP002553 LOCUS 297816 bp DNA BCT 07-MAR-2001
      DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 4/20.
      ACCSSION AP002553 BA000007
      VERSION AP002553.1 GI:13360211
      KEYWORDS
      SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
      DNA.
      Escherichia coli O157:H7
      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
      Escherichia.
      REFERENCE
      1 (sites)
      Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
      Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
      Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
      Sasakawa,C. and Shinagawa,H.
      Complete nucleotide sequence of the prophage VT2-Sakai carrying the
      verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
      derived from the Sakai outbreak

      JOURNAL
      MEDLINE
      REFERENCE
      AUTHORS
      Genes Genet. Syst. 74 (5), 227-239 (1999)
      20198780
      2 (sites)
      Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
      Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
      Hayashi,T.
      Comparative analysis of the whole set of rRNA operons between an
      enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
      Escherichia coli K-12 strain MG1655
      Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
      20557356
      TITLE
      JOURNAL
      MEDLINE
      REFERENCE
      AUTHORS
      3 (sites)
      Yokoyama,K., Makino,K., Kubota,Y., Matanabe,M., Kimura,S.,
      Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
      Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
      Shinagawa,H.
      Complete nucleotide sequence of the prophage VT1-Sakai carrying the
      Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
      O157:H7 strain derived from the Sakai outbreak
      Gene 258 (1-2), 127-139 (2000)
      20564182
      TITLE
      JOURNAL
      MEDLINE
      REFERENCE
      AUTHORS
      4 (sites)
      Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
      Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
      Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
      Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
      Shinagawa,H.
      Complete genome sequence of enterohemorrhagic Escherichia coli
      O157:H7 and genomic comparison with a laboratory strain K-12
      DNA Res. 8 (1), 11-22 (2001)
      21156231
      TITLE
      JOURNAL
      MEDLINE
      REFERENCE
      AUTHORS
      5 (bases 1 to 297816)
      Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
      Hayashi,T.
      Direct Submission
      Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
      Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
      Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
      URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
      Fax:81-6-6879-2047)
      genome project.
      COMMENT
      FEATURES
      LOCATION/Qualifiers
      source
      1..297816
      /organism="Escherichia coli O157:H7"
      /strain="O157:H7"
      /sub_strain="RIMD 0509952"
      /db_xref="taxon:83334"
      79..1245
      /gene="ECs0753"
      79..1245
      /gene="ECs0753"
      /note="ECs0753"
      /note="Similar to SUCC_ECOLI g11786948 percent identity
      100 in 388 aa (Conserved in E.coli K-12)"
      /codon_start=1
      /transl_table=11
      /evidence=not experimental
      /product="succinyl-CoA synthetase beta subunit"
      /protein_id="BAB34176.1"
      /db_xref="GI:13360212"
      /translation="MNLHEYQAKOLFARYGLPAPVGYACTTPREAEASAKIGAPVY
      VKCOVHAGGRKAGVGVVNSKEDIRAFANMIGKRYVTQTPANCOVPOITVEAT
      DIKEILGAVYDSSRRVYFMATBEGVEIEKVAETPPLHKKVALDPTGMPYOG
      RELAFKLGDEKLVQDFKIFMGATLFLERDLALLETINPLVITKQGLDLCGKIGA
      DGNALFROPDLREMRQSDPPEAQAOEWELNVALDGNITGVNAGAGLAMGTMDIV
      KLHGEPANFLDVGAGATKEBVEAFKIIISDKVKAVLVNIFGIVRCDLIADGIIG
      AVAEVGNVPVYVLEGNNAELGAKKILADGLNIIAAKGLTDAQGVAAVEEK"
      1245..2114
      /gene="ECs0754"
      1245..2114
      /gene="ECs0754"
      /note="ECs0754"
      /note="Similar to SUCC_ECOLI g11786949 percent identity
      100 in 289 aa (Conserved in E.coli K-12)"
      /codon_start=1

```

```

gene
CDS
    complement(6092..7372)
        /gene="ECS0758"
        /complement(6092..7372)
            /note="probable transport protein,similar to glutamate/aspartate transport proteins (proton glutamate symport proteins) e.g. [Bacillus stearothermophilus] aa,also similar to C4-dicarboxylate transpor"
                /codon_start=1
                /transl_table=11
                /evidence-not_experimental
                /product="putative transport protein"
                /protein_id="BAB34181.1"
                /db_xref="GI:13360217"
                    /translation="MKKISLTMTLLALVIGMIIGVVLNNTASPEPAKIYAQEIIFITIELLIMIIAPLVSTLVLVGIAKKMDAKAIRITSKLPLEFICASSLSIALGITYNFEPMGTGINVAHGAETTGVAAEPTLVLPISHAFPTSIVDAMAHNEILDIVESFLGCSLTAVIKERGSAILVHALDSLAAHMLKLTGVYMLEADPLTYFAAISAIARGIAYVASAGLFEGEEFYFTMLLIALLGLAIYYVPCIRRTLRALSEPALLAFTSSSEBAATGFLEKEOFGFSVPKIASFPVLPIGVSNLVOSMAYCGEPARTAOACNHILTSGEDITPLLILUTSKGNACYPRASMYVIATLINOFNPINAGILLMGDPFLDMKRSAITNMMSAGAAMAWSRMEGEHEHGECRGKALKTPNESVALP"
                        complement(7533..7850)
                            /gene="ECS0759"
                            /complement(7533..7850)
                                /gene="ECS0759"
                                    /note="unknown"
                                        /codon_start=1
                                        /transl_table=11
                                            /evidence-not_experimental
                                                /product="hypothetical protein"
                                                    /protein_id="BAB34182.1"

Query Match      100.0%; Score 25; DB 1; Length 297816;
Best Local Similarity 100.0%; Pred. No. 1;.9e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY       1 gcaattgcagtcctcgtcacgqtgcag 25
          |||
Db 92409 GCATTTCGATCTCCGTCAAGTGCAG 92433

RESULT 15
ECOBIOD   128 bp     DNA         BCT           03-JAN-1995
DEFINITION Escherichia coli b1obn (bioB) gene, early terminator region.
ACCESSION M27731
VERSION   M27731.1 GI:341755
KEYWORDS bioB gene; biotin.
SOURCE    Escherichia coli (strain K-12) DNA.
ORGANISM Escherichia coli
REFERENCE 1 (bases 1 to 128)
AUTHORS   Nath,S.K.
TITLE     Attenuation of transcription of biotin genes in Escherichia coli
JOURNAL   Can. J. Microbiol. 34 (12), 1288-1296 (1988)
MEDLINE   89167942
FEATURES
Source
    Location/Qualifiers
        1..128
            /organism="Escherichia coli"
            /strain="K-12"
            /db_xref="taxon:562"
        4..128
            /gene="bioB"
        4..9
            /gene="bioB"
        20..>128
            /gene="bioB"

```

```

terminator      103..121
                /gene="b10b"
                /note="early terminator"
BASE COUNT      31 a      28 c      34 g      35 t
ORIGIN
Query Match      84.0%; Score 21; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.0061;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 gcaatcgatccctgcagagt 21
        |||
        59 GCATTTCGATCCTCCTCAGGT 79

RESULT 16
AF250776      5526 bp      DNA      BCT      31-JAN-2001
LOCUS
DEFINITION      uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in
                  modC-b10a intergenic region, DAPA-aminotransferase BioA (bioA),
                  biotin synthase BioB (bioB), KAPA synthetase BioC (bioC), and
                  biotin biosynthesis protein BioC (bioC) genes, complete cds; and
                  dechlorobiotin synthetase BioD (bioD) gene, partial cds.
ACCESSION      AF250776
VERSION
KEYWORDS
SOURCE
ORGANISM      uncultured bacterium pCosHE2.
                  Bacteria; environmental samples.
REFERENCE
AUTHORS      1 (bases 1 to 5526)
                  Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streif, W.R.
                  Direct cloning from enrichment cultures, a reliable strategy for
                  isolation of complete operons and genes from microbial consortia
                  Appl. Environ. Microbiol. 67 (1), 85-99 (2001)
TITLE
JOURNAL
MEDLINE      20575196
PUBMED      11133432
REFERENCE      2 (bases 1 to 5526)
                  Entcheva, P., Liebl, W. and Streif, W.R.
                  Direct Submission
AUTHORS
TITLE
JOURNAL
FEATURES
source
                  Location/Qualifiers
                  1..5526
                  /organism="uncultured bacterium pCosHE2"
                  /db_xref="taxon:143797"
                  /clone="pCosHE2"
                  /note="unknown organism, cosmid clone derived from
                  environmental consortium"
                  complement(52..528)
                  /note="ORF1"
                  /codon_start=1
                  /transl_table=11
                  /product="hypothetical 17.1 kDa protein in modC-b10a
                  intergenic region"
                  /protein_id="AAG60577.1"
                  /db_xref="GI:12620125"
                  /translation="MKLISNDLRDQGLPHRHEGNGYDGNISPHLAMDVDPAGTK
                  SPVYCYDPDAPGSGMHWVYNLPADRVLPQGGSGIYAMPDGVLTQTRIDFGAG
                  YDGAAPKGGTHYITFYVHALDIERLDIVDEGASGANGVFNHSLASASTAMFS"
                  complement(587..1876)
                  /gene="bioA"
                  complement(587..1876)
                  /gene="bioA"
                  /note="7.8-diaminopelargonic acid
                  synthetase-aminotransferase"
                  /codon_start=1
                  /transl_table=11
                  /product="DAPA-aminotransferase BioA"
                  /protein_id="AAG60578.1"
                  /db_xref="GI:12620126"
                  /translation="MTTDDLAFDQRIHWHPTSMISPLVYVVAEGCELLISDGRK
                  LVDSGSMWMAIHGYNHPQLNAAMKSOIDAMSHVMFGGITHAVATLECRKLVAMTRPP

```

```

1KCVFLADSSGSAVEAVAMKMLQYMAKGEARORPTFRNGVHGDTEGAMSCDDPDS
MISLWKGYLPENLPAPAPQSRDQSDMDRDWNGRRLMAAHRHEIAVITIPPYOGAG
GKRWYHPWELKRIKICDREGILLIADITAGFGTGKLFACERAEIAPDLICGKAL
TGGMTLTSATLITREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAIIESDWO
QVADI EVOILREQLAPARDAEMVADRVYGAIGVETTPPVMAALOKFEVQGVIRP
FGKLIYLMPPYITLPOLQRLTAANRAVQDETEFCQ"
1963..3003
/gene="bioB"
1963..3003
/gene="bioB"
1963..3003
/gene="bioB"
/transl_table=11
/product="biotin synthase BioB"
/protein_id="AAG60579.1"
/db_xref="GI:12620127"
/translation="MAHRPMTLSQVTELFKRLDILFEAQYVROHPDPROVYST
LISIKTGAPEDCKYCPQSSRYKTGLEAEKLEMEVOYLESAKAKAGSTRFCGGAAM
KNPNERDMPYLEOMVQVYKDLGLACMKTGLTSEQAQRLANGLDYNNHLDTSPEF
YGNLITTRTYOERLDTLEKVRDAGIKVSGGIVGIGETVRAGILLQLANLPPTPS
VPINMLVYVKGTPIADNDVDVAFDFIRTIIVARIIMPTSVYRLSAGRQNNRQOAMC
FMAGANSIFGCKLLTTNPNBEDKDLQFLKGLGIMPQOTAVIAGDNEQQRLEQALMT
PPIDEYVAAAL"
3000..4154
/gene="bioF"
3000..4154
/gene="bioF"
/note="8-amino-7-oxononanoate synthase"
/codon_start=1
/transl_table=11
/product="KAPA synthetase BioF"
/protein_id="AAG60580.1"
/db_xref="GI:12620128"
/translation="MSMOEKIINAIDARADALRRYPVAGACGRILVADPROYLNF
SSNDLGLSHRPQITRAMQGAEBOPCTGGSGSHYSGYSVYHQLEEBLEAMLTSTRA
LITISGFANQAVIAAMAKEDRIADRLASLISLEASLSPQLRRAHVDTHLAR
ILASPGQQLVYTEGVFSMDGDSAPLAEIQVYQOHNQWMLMDVADGATYIGQGG
SCMLOKVPBELVTEFGKFGVSGAAVCSSTVADYLIQFAHRLIYSTMPAPQAOL
RASLAVIRREDGARREKLVSIIAHFRAGVODLPFTLADSCSAIPLITVGNSSALQL
AEKIQGCCWVTAIRPPVPGATRLRLITLAAHEMQDIDRLLEVLNGNC"
4141..4896
/gene="bioC"
4141..4896
/gene="bioC"
/note="reaction step prior to synthesis of pimeloyl-CoA"
/codon_start=1
/transl_table=11
/product="biotin biosynthesis protein BioC"
/protein_id="AAG60581.1"
/db_xref="GI:12620129"
/translation="MATVAKOAIAAAFGRAAALYEOHADLOROSADVLLAMLPKRYT
HYLDAGCGPWTRHMRERHAQVYALDLPMLVOARQKADAHYLAGDIESPLATA
TFDLAWSNLAVOMCGNSTALRELYRVKSGVAVFTIYNGSLPELHOAMQADRRP
HANRFLPPDEITEOSLNCVYHQHITPITLIMRDDALSAMRSIKGIGATHLHEGRPRIL
TRSQLRQRLAMPQOQGRYPLTHLFLGVARE"
4883..5526
/gene="bioD"
4883..5526
/gene="bioD"
/note="DTB synthetase"
/codon_start=1
/transl_table=11
/product="dechlorobiotin synthetase BioD"
/protein_id="AAG60582.1"
/db_xref="GI:12620130"
/translation="MLVSKRYFVNGTQTEVNGKYTASCALLDQAAKAAGYRTAGYPPVAS
GSEKTPEDLRNSDALAIORNSSIDLDATVAVPYFAETSPHITISABGRTISSVMS
SGLRALTEQADWVLEVGAGVFTPLSDTFEFDVWQDEQLEPVILIVGKLCQCNHAAHL
TAAQIOHAGLQIAGVANDVYPPGKRHAKEYITTLTRMI PAVLLGEIEMLAENBP"
BASE COUNT      1274 a      1507 c      1567 g      1178 t
Query Match      84.0%; Score 21; DB 1; Length 5526;

```

Best Local Similarity 100.0%; Pred No. 0.0066;
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatctgacccctgcaggt 21
|||||
db 2061 GCATTTCGATCCGCGCAGG 2081

RESULT 17
MTC1429B 2980 bp DNA BCT 03-AUG-2001
LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 38/162.
DEFINITION 296797 AL123456
ACCESSION 296797.1 GI:3261811
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
1 (bases 1 to 2980)
Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,
Rekai, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squires, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and
Barrell, B.G.

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 2980)
Parkhill, J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams: Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:2190471.

COMMENT

Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in TParse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.

FEATURES
source
1. .2980
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
misc_feature
1. .613
/note="fragment designated v042. Does not represent a
physical clone"
misc_feature
39. .42
/note="inverted repeat at 3'-end of IS6110"
misc_feature
51. .1142
/note="4 bp duplication of IS6110 target sequence"
repeat_unit
/gene="Rv0797"

gene
CDS

/note="IS1547"
51. .1145
/gene="Rv0797"
51. .1145
/gene="Rv0797"
/note="Rv0797, (MTC1429B.03c-MTV042.07), transposase for
IS1547, len 364 aa; almost identical to (but 20 aa shorter
than) gp1Y13470|MY13470.2 Mycobacterium tuberculosis
gene (383 aa). Also similar to other transposases eg
MAIS110A_1 M. avium insertion sequence Q48909 transposase
(464 aa), fasta scores: opt: 226 z-score: 275.7 E(1):
2.4e-08, 30.7% identity in 199 aa overlap. Also slight
similarity to M. tuberculosis protein MTCY39.03c (24.8%
identity in 141 aa overlap)"

source

/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"

gene

/product="hypothetical protein Rv0797"
/protein_id="CAB09573.1"
/db_xref="GI:3261812"
/translation="MNVVGTDAHKYSHFVATDEVGRGLEGKTKATTAAGATAIMNA
RQFCLELHIEDCRNMSARLEKRLDLAGQVAVRVPRKLAQKRSKRSKSDPD
ALAVARAVRETDLPATHDERSRELKLTORRDVLAQKRSALNRLMHLHEIDPER
APAAASRLAARHQALRTMLDTPGLVLAELRAELTDIIRLGETINTLAQRI SARVH
VAPALLEIPGCAELTAARIVGEAAVGVTRFKSAFAAACHAAVAAPVPGSGTAGQMRUS
RSGNOLNALRIALTOIRMTDSRGOAYVORLODAGKTRKAALRLKRLARIVFOA
LRTVQPSSEHNPAAACHRSYCSKSLSG"
513. .2334

CDS

/complement(1135. .1932)
/gene="Rv0798c"
/note="Rv0798c, (MTC1429B.02), len: 265 aa; 29 kDa
antigen: cfp29, identical to gp1Y12820|MY12820.1
Mycobacterium tuberculosis cfp29 (265 aa), 99.6% identity
in 265 aa overlap. Also highly similar Q45296 linoclin M18
from BULNM18P_1 B. linnens; 58.5% identity in 265 aa
overlap"

gene

/product="hypothetical protein Rv0798c"
/protein_id="CAB09572.1"
/db_xref="GI:2190473"
/translation="MNLYRDLAPVTEAMAELEAATFEKRLAGRRVVDSPDG
PVTAASVGRLELDVAPRTGVIARASPELVRLVPTLPSRNEIDVDERSKSDME
PVKAARKLAPEFDRITFEFGYSAASIEGRSSAPALIPEDPREIPDIVISALSEL
RLAGVGPYSVLISADVYKVSSTSDHGPRIEHLNRIYVDGIINAPAIIDGAFVLTFR
GGDFDQIDTDAIGASHTPTVRLYIDETLITFLCYTRASVALSH"
complement(1929. .2936)
/gene="Rv0799c"
/note="Rv0799c, (MTCY0797A.10-MTC1429B.01), len: 335.
similar to ECAP00330.8 E. coli K-12 p76536 (308 aa) E(1):
2.2e-30, 37.4% identity in 297 aa overlap. Also similar to
M. leprae, Q50021, U2266C, (146 aa), fasta scores, opt:
147 z-score: 188.8 E(1): 0.0016, (33.3% identity in 117 aa
overlap) and Q50020, U2266B, (27 aa), fasta scores, opt:
94. E(1): 1.3, (56.5% identity in 23 aa overlap)"

gene
CDS

/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0799c"
/protein_id="CAB09574.1"
/db_xref="GI:3261813"
/db_xref="SPTREMBL:007180"
/translation="MAVPAVSPQPLIATPTPAIFLVATIGADGATVADALSKISGL
VRAIGFRDPTKHLISVYSIGSDAWRLPAGRPTEHLHPVELTGRHRAPARCGLLE
HRAETMDVCFELAGRIKSKGDAVTVVDEHGFREFRDLLGFEVDGTENSGPIAI

KATTIGEDENFAGSCYVHVOKYVHDMASMESLSTVEOEERVIIGRTKLDELIDNMRP
ANSHVNLNITDDGTERKIVRNHMPGEGKGYGYFICYSRTPTVIEQMLRNMTI
CDPAGINTDRULDPSTAVTGGLFFSPPTIDFLDHPPLPQAATPILAGSLISGLKGS
R"

RBS
comp1ement(1941. .1945)
/gene="Rv0799c"
/note="possible RBS upstream of Rv0798c"
2331. .22980
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="Y7H7A"

BASE COUNT 566 a 965 c 939 g 510 t
ORIGIN

Query Match 64.0%; Score 16; DB 1; Length 2980;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 attcgatccctgca 18
|||||
Db 2424 ATTGATCCTGCTCA 2439

RESULT 18
AE006972 14756 bp DNA BCT 27-APR-2001
LOCUS Mycobacterium tuberculosis CDC1551, section 58 of 280 of the
DEFINITION complete genome.
ACCESSION AE006972 AE000516
VERSION AE006972.1 GI:13880364
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis CDC1551.
Mycobacterium tuberculosis CDC1551.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE
AUTHORS
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Gill, J., Mikula, A., Utterback, T., Weidman, J., Khouri, H.,
Salzberg, S.L., Delcher, A., and Bishai, W.
Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
Unpublished
2 (bases 1 to 14756)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., Deboy, R., Dodson, R., Gwin, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
Direct Submission
Submitted (25-APR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. 14756
/organism="Mycobacterium tuberculosis CDC1551"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="clinical strain"
196. .1347
/gene="MT0818"
196. .1347
/gene="MT0818"
/note="similar to GB:M31716.SP:PI9834.PID:551996;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="IS1547, transposase"
/protein_id="AAK45060.1"
/db_xref="GI:13880365"

/translation="MCPPTGPTSTPPVKEATTWVYCTDAHKYSHTFVATDEYRGL
GKTKATTAGATATAIMARQDFGLTWGIEDCRNNSARLERDLAAGQOVAVPRK
LMAOTRRKARSRKSDPIDALAVARAREDEPLATHEDETSRELKLTORBDLVQ
RNSATNRLRWLVHEIDPEPARAPARLDAKHOQALRLTIDPQGLVAELTADLI
RLTGEINTLTAORISARHVOVAPALLEIYKQALVETLQVGAACVTPKRSAAACGA
AVAPVWNGNTAGOMRUSRSGNRLNMAHLRLTORMDPSGQAYVYORLQDAGKT
KRAALRCLKRLIARTVFOALRTVYHDPSESHOAPAACHRSICSHSLSG"

gene
complement(1337. .2134)
/gene="MT0819"
complement(1337. .2134)
/gene="MT0819"
/note="similar to GP:2231022; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="bacteriophage CP29"
/protein_id="AAK45061.1"
/db_xref="GI:13880366"

CDS
/translation="MNNLRDLAVYTEANAELEAARFKRHIAGRVYVDSDECG
PYTAASVIGRLIDVKAPTNGVIAHLRAKAPLVRLRPPTLSRNEIDVGRSKDSME
PVKEAKRLAFVEDRTIEGYSAASIEGIRSSANPALTLPEDEPREIDPVISQALSEL
RLAGVDGPPSYLSADVYTKVSETSDHGYPIREHLNRLVDGDIIMADIGAEVLTTR
GGDFDLOGTVAIGYASHDPDVTFRLYIQETLFTCYTAASVALSH"
complement(2131. .3138)
/gene="MT0820"
complement(2131. .3138)
/gene="MT0820"
/note="identified by Glimmer2; putative"
/transl_table=11
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="AAK45062.1"
/db_xref="GI:13880367"

gene
complement(1337. .2134)
/gene="MT0821"
complement(1337. .2134)
/gene="MT0821"
/note="similar to GP:5457236; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="aminopeptidase, putative"
/protein_id="AAK45063.1"
/db_xref="GI:13880368"

CDS
/translation="MAATFAGLCEFFIDASPPPHVCATVAGRLLGAGYRELRAWRMP
DKPGRYFTVRASGLVANNASQSGHTVOVFRIVGAHDSPLRVKQHPDLRVAQMVVA
LQPYGCVLHSMILDRDLGISRLSVRGSTVSRILVLDPLRLVQLATHLAEDKRS
LTLDPPRIHNAWVGSEVEFVYGVAVGVAADVLADMTHTDTPALGASVN
GTALLSPRIADNDONASCYAGMEALLADVNSAGSPVATITPDHEVGASRCHGS
DLSSYLERIVLAAGTRDEFLRLRTTSMASDMATHTPNTPDHRESHPLEVAVG
PVKAPVRLNRTATDGRTAARALACQAGVPMQRYEHRADPCGSTTGPLAARTGIP
TVDVGAOGLMHASRELTMGAHDVAAYSAALQALSAELSEA"
complement(4837. .5493)
/gene="MT0822"
complement(4837. .5493)
/gene="MT0822"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK45064.1"
/db_xref="GI:13880369"

gene
complement(1337. .2134)
/gene="MT0823"
complement(1337. .2134)
/gene="MT0823"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="IS1547, transposase"
/protein_id="AAK45065.1"
/db_xref="GI:13880370"


```

gene
5685..7949
/gene="MT0823"
5685..7949
/gene="MT0823"
/note="similar to PID:699222 PID:2076635; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="phosphoribosylformylglycinamide synthase 11"
/protein_id="AAK45065.1"
/db_xref="GI:13880370"
/transl_table="MLDIVEHAATPPDOPPYGELGKDEYRIRIQLGRPDTL
AMYSVMSESEKSKSVHLRYGFTSTSDPMAMAGIGENAGVDDIDGNAVPKRY
ESHNHSTVEPYOGATGCGTIVRDIMAGCARVAVMDLRKADAPPTORVLDVY
RIGIGYNSGLPNIIGETVEDPCYAGNPLVNAICVGLKREKHLAFASAGNKTLL
FGAR7GLDGIIGVSVLASDPEVAEGSRKLIPIVOYDVPPEKYLICCELEIYAGIYI
GIODLGAGLSCATSEIASAGDGMTIOLDSVPLAKEMTPAVILCSESEKMCAYVS
PKNVDAIYAVCRKWEVLAIVIGETDGDRIQIWHGETVVDVPPRTVAHEGPPYQRPV
ARPDIDALNADRSKLSRPVGTDELRAITLLGSPHLCRAFTIEOYDRYRGNTV
LAERADGMLRIDSOTRGIAVSTDSAGRTLLDPAAGNLAFAVRNVAICAPVY
AVTNCLNFGSPEDGVMMOTQAVRGLADGADLGIPTGVGNVSTNOTGSAALIPTP
VGVGLYIDVYRRRIPTGLGAEPEGTMLLGDTRDEPDGSMVAQVADHLGGLPVVD
LAREKLAVALSSASRDGLSAADHISEGLAQAIYESALAGETGGRIVLPEGADPEV
LRFESAGRVAVVPRTEESRFRMCERAGLPVIRIGVYDGSDAVEOGLFAVSLAE
LRATSEAVLRPRYG"
7946..8575
/gene="MT0824"
7946..8575
/gene="MT0824"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45066.1"
/db_xref="GI:13880371"
/transl_table="MSRLRALSLAAGVGSLSVPLPAPMRIPLDGLSGSVLVTR
ATMGLPRLIAGLRICMAGAAATAIATTPVPMVRLSMSRELPSAPVPMVIMHI
PGGVNAEAPAFKALATIGARAGRGGRILDAGACGISHIDARATGSPVLVTVA
TGIGMFGMLADRSSLSAALPLTHLAINEGAAVAVLVRRCISITRL"
8696..9652
/gene="MT0825"
8696..9652
/gene="MT0825"
/note="similar to PID:699231 PID:2076636; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="icc protein"
/protein_id="AAK45067.1"
/db_xref="GI:13880372"
/transl_table="MHRRAAEHPPRYVILHISDTHJGCDRLYCAVADDRGLGL
LEQUNQGLRDAIVFTCDLADKGEPAAYRRLGLVEPFAQLAEALVWAGNHDDPA
ELRFLDEAPASMAPIIDRVCMIDGLRIIVLDTSPVGHNGEIRASQGLWAEENBDA
PDGTIALHHPPIPSVILMAVTVLELDOALRGVLRGTDAIILAGLHSTNATFVG
IPVAVASATCTODLTVAGGTGRGDAKCCNLVHVYDPIVYHVIPLGGGETVGTFF
SPGQARKRIASGTFLEPSRSDSLFKHPPMVLTSSARSPVD"
complement(9597..11195)
/gene="MT0826"
complement(9597..11195)
/gene="MT0826"
/note="similar to PID:699232 PID:2076637; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cpst protein"
/protein_id="AAK45068.1"
/db_xref="GI:13880373"
/transl_table="MPKISSRDGGRPAROTVNPITIVRRCKIARLESCTPQEQIJD
LVFERKVLNADIPYLLIRNKNRPVLAINIELRAGLRRAAACATBEMPAKTIDEP
GLSPVLAFTDGLSQLVDRPVYVLRRIYAPGFRGPGFVGLQFWEYEEVYICPVE
NLSRKVLPNEITPTNKLKYKWPITLGDGAFAPASADVFDIDMVSWMVDSGSPER
ARRMAQSYVVGEGDDAEARIQIDELKVALRSVMFAPWIRIRIFATDSTPPMILA

```

```

EHPKTIIVRAEDHESDRSALPTYNASHAVESQLHHIIPGLSEHFLYSNDMEFGRPKAS
MEFSGVTRFTEAKTRIGLGNANPARSGFENARVARNOLLEDFRGVITRHEHTAV
PLRKSULTEMREPEEPFAPRTAASPRSDNDSVYNSPYHYVALMTGRAVPOEKAYVL
YVDTTSYAGLRILPKLRKHNGYDFCLNDGSPFPEYPAARERVRVSTLENTFPIPAW
EKIADVSRRDPDVPRTSAPSDGA
complement(11230..11337)
/gene="MT0827"
complement(11230..11337)

Query Match      64.0%; Score 16; DB 1; Length 14756;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      3      attcgatcctgcgtca 18
Db      2626 atttcgactctgcgtca 2641

RESULT 19
MSG8136S
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mycobacterium leprae (clone: cosmid B13) (tissue library: lorist 6)
DNA.
Mycobacterium leprae
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
Mycobacterium.
1 (sites)
Bijlmeester,K., Honore,N., Woods,S.A., Caudron,B. and Cole,S.T.
use of an ordered cosmid library to deduce the genomic organization
of Mycobacterium leprae
Mol. Microbiol. 7 (2), 197-206 (1993)
93188700
2 (bases 1 to 42923)
Smith,D.R., Richerich,P., Rubenfield,M., Butler,C., Lee,H.-M.,
Xu,Q., Gunderson,K., Chung,M., Maher,J.K., Delonghery,C.,
Aldrich,T., Mitch,J., Tulig,C., Smyth,A., Drill,S., Avrich,A.S.,
Rice,P., Abendschan,K., Aldredge,T., Delonghery,C., Kirst,S.,
Saffer,H., Connolly,S., McDougall,S., Bijlmeester,K., Bergh,S.,
Cole,S., Robison,K., Jaehn,L., Gryan,G., Church,G.M. and Mao,J.
Prepublication submission
Unpublished (1996)
This sequence was generated by the Genome Sequencing Center at
Genome Therapeutics Corporation (Collaborative Research Division),
100
Beaver St., Maltham, MA, 02154. Please contact Doug Smith
(smith@ctic.com) for further information. The sequence represents
the
insert of a lorist 6 cosmid clone from a mapped set of clones
constructed from M. leprae genomic DNA isolated from armadillo
liver
[3]. The sequence may not represent the entire cloned insert of
the
cosmid if an overlapping region was previously sequenced from
another
clone. Coding sequences larger than 60 amino acids were predicted
on
the basis of codon usage and homology information. An attempt was
made
to locate the most probable start site based on codon usage,
homology,
the presence of a Shine-Dalgarno sequence, or overlapping orf that
suggested translational coupling. It is possible that the actual
start
site differs from the one selected.
location/Qualifiers
1..42923
/organism="Mycobacterium leprae"
/specific_host="Dasyus novemcinctus"

```

FEATURES

source

```

/db.xref="taxon:1769"
/clone="cosmid B13"
/tissue="Liver of the armadillo was used to isolate the
Mycobacterium leprae."
BASE COUNT      9493 a 13528 c 11679 g      8223 t
ORIGIN

Query Match      64.0%; Score 16; DB 1; Length 42923;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 ttccatctctcagcag 20
|||||
Db 20721 TTCGATCCTCGTCAGC 20736

RESULT 20
MCL458/c 43839 bp DNA BCT 27-AUG-1999
LOCUS Mycobacterium leprae cosmid L458.
ACCESSION AL049478
VERSION AL049478.1 GI:4539121
KEYWORDS acps; CoA:apo-[Acyl Carrier Protein]; DNA polymerase III alpha
chain; dnaE; fas; ilc; ilvA; isoleucyl-tRNA synthetase;
L-asparaginase; L-glutaminase; lipoprotein signal peptidase; lspa;
oxidoreductase; pseudogene; pseudouridine synthase; Rbp element;
rluc; theonine dehydratase biosynthetic; type I fatty acid
synthase.
SOURCE Mycobacterium leprae.
ORGANISM Mycobacterium leprae
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
Mycobacterium.
REFERENCE 1 (bases 1 to 43839)
AUTHORS Eiglmeier, K., Honore, N., Woods, S.A., Caudron, B. and Cole, S.T.
TITLE Use of an ordered cosmid library to deduce the genomic organization
of Mycobacterium leprae
JOURNAL Mol. Microbiol. 7 (2), 197-206 (1993)
MEDLINE 93188700
REFERENCE 2 (bases 1 to 43839)
AUTHORS Seeger, K.J. and Harris, D.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 43839)
AUTHORS James, K.D., Parkhill, J., Barrell, B.G. and Randeram, M.A.
JOURNAL Direct Submission
TITLE Submitted (15-MAR-1998) Mycobacterium leprae sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire,
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Dr.
Stewart T. Cole, [3] Unite de Genetique Moleculaire Bacterienne,
Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15,
France Requests for cosmids should be sent to Karin Eiglmeier
(keiglmeier@pasteur.fr)
Notes:
The Sanger Centre is funded to complete the sequence of M. leprae
by the Helser Program for Research in Leprosy and Tuberculosis of
The New York Community Trust.
Work in Paris is supported by the Helser Trust, the Association
Francaise Raoul Follereau and the Groupement de Recherches et des
Etudes des Genomes (GIP-GREG).
Details of M. leprae sequencing at the Sanger Centre are available
on the World Wide Web.
(URL, http://www.sanger.ac.uk/projects/)
CDS are numbered using the following system eg MLCB458.01c. MLC (M.
leprae), cB33 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given

```

FEATURES

source

where these have been used to deduce the initiation codon. All CDS over 100 codons have been analysed. Gene prediction is based on positional base preference in codons especially where there is an increase in the observed/expected third position G + C. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Location/Qualifiers

1. 43839

/organism="Mycobacterium leprae"

/db.xref="taxon:1769"

/clone="cosmid L458"

1. 569 /note="overlap with cosmid B1701 from 19932 to 20500"

complement(97..531)

/gene="MLCB485.01c"

complement(97..531)

/gene="MLCB485.01c"

/note="MLCB485.01c, hypothetical protein, len: 114 aa;

unknown function, possible CDS based on amino acid

composition and frame analysis"

/codon_start=1

/transl_table=11

/label="MLCB485.01c"

/product="hypothetical protein MLCB485.01c"

/protein_id="CAB39566.1"

/db.xref="GI:4539121"

/translation="MQSLRRHRCSTFFHYGYSILPAORAVVRTSTADTLSDTMTGA

AGENGRVTHQAPAVDOVGRSGSVTVTPAPRSVGRVSPADLDDSFCHSGTNPFLD

QVAVFPSSADLKPOLNTAKFRFVADHGVEQWIDFVAMVTP"

complement(827..1210)

/gene="MLCB458.02c"

complement(827..1210)

/gene="MLCB458.02c"

/note="MLCB458.02c, hypothetical protein, len: 127 aa;

unknown function, N-terminus similar to internal sequence

of YP22_MYCTU (EMBL:Z80226) RW0778 (MTCY369.22)

M.tuberculosis hypothetical protein (414 aa), fasta

scores: opt: 139 z-score: 189.2 E(): 0.0031, 31.48

identity in 102 aa overlap"

/codon_start=1

/transl_table=11

/label="MLCB458.02c"

/product="hypothetical protein MLCB458.02c"

/protein_id="CAB39567.1"

/db.xref="GI:4539123"

/translation="MLEKNRPQIFDSDRILKASPEAWLSRKEIDLSTIFEYTLK

LTAERKNPSDDIWRILASAVITGNDKPPRLPANELFFVLTGNSNATKHTTYW

TAGVHHESRPDKTTPRSIGSIAPQR"

complement(1581..1850)

/gene="MLCB458.03c"

complement(1581..1850)

/gene="MLCB458.03c"

/note="MLCB458.03c, hypothetical protein, len: 89 aa;

unknown function, improbable CDS based on frame analysis"

/codon_start=1

/transl_table=11

/label="MLCB458.03c"

/product="hypothetical protein MLCB458.03c"

/protein_id="CAB39568.1"

/db.xref="GI:4539124"

/translation="MQLAPEMRGGEPVFTIAKVYTIILSGHONGRTSSHREYSL

FLTHRHILCLSSSDGVCGKSDDEPRRPVRSQNTGSELPF"

complement(1841..2110)

/gene="MLCB458.04c"

complement(1841..2110)

/gene="MLCB458.04c"

```

/note="MLCB458.04c, hypothetical protein, len: 89 aa;
unknown function, possible CDS based on amino acid
composition and frame analysis"
/codon_start=1
/transl_table=1
/label="MLCB458.04c"
/product="hypothetical protein MLCB458.04c"
/protein_id="CAB39569.1"
/db_xref="GI:4539125"
/translation="MFERMTAWRIKPIKIAAREIPATLVNSVNPRIPLMPASYRRM
RCDPTSTRVDSEDSFVIDSHQSLIVAKRSQIAKQIAISGCS"
3625..4344
/gene="MLCB458.05"
3625..4344
/gene="MLCB458.05"
/note="MLCB458.05, possible secreted protein, len: 239 aa;
unknown function, similar to TR:P95028 (EMBL:283863)
RV2525c (MTCY159.31) M.tuberculosis hypothetical protein
(240 aa), fasta scores: opt: 1358 z-score: 1535.9 E(): 0,
82.2% identity in 241 aa overlap. Contains possible
N-terminal signal sequence"
/codon_start=1
/transl_table=1
/label="MLCB458.05"
/product="hypothetical protein MLCB458.05"
/protein_id="CAB39570.1"
/db_xref="GI:4539126"
/translation="MSYSRDVLFKATVTPGGLGCGAAALCAVASTAGSITLID
YAAGVIPASQIRATGAVGAIKRVSDPGTAVKCPDVOTARDIINNGLTVCYVG
KGNATVMDIGATAGLRHAGRGVQLHTAAGSPVAPYASIDSPTEYQKQVAPYLR
SMESVIGHORIGVANSRTIAMALQDLASYFMQHMWSPKGYTHPANLHOVIDBR
TVGGVGVNRTIKKPCFGQWA"
4956..14186
/gene="fas"
4956..14186
/gene="fas"
/note="MLCB458.06, fas, probable type I fatty acid
synthase, len: 3076 aa; similar to many eubacterial and
eukaryotic FAS e.g. TR:O59497 (EMBL:X87822) Brevibacterium
ammonigenes FAS (3063 aa), fasta scores: opt: 5361
z-score: 5527.0 E(): 0, 44.3% identity in 3106 aa overlap.
N-terminus shows similarity to FAS1_CANAL (EMBL:X74952)
Candida albicans FAS beta subunit (2037 aa) (27.3%
identity in 1928 aa overlap) and C-terminus to FAS2_CANAL
(EMBL:L29063) Candida albicans FAS alpha subunit (1885 aa)
(27.1% identity in 1739 aa overlap). C-terminus similar to
N-terminus of TR:O69474 (EMBL:AL023635) MLCB1243.20c,
Kasa, M.leprae beta-ketoacyl-ACP synthase (414 aa) (27.5%
identity in 386 aa overlap) and TR:O69473 (EMBL:AL023635)
MLCB1243.19c, Kasa, M.leprae beta-ketoacyl-ACP synthase
(425 aa) (25.1% identity in 382 aa overlap). Equivalent to
TR:P95029 (EMBL:283863) RV2524c (MTCY159.32-MTY009.09c)
M.tuberculosis FAS (3069 aa) (85.8% identity in 3081 aa
overlap). Contains Pfam match to entry PF00698
Acyl-transf. Acyl transferase domain, score 18..20, E-value
1.9e-08, PS00017 ATP/GTP-binding site motif A (P-loop) and
PS00606 Beta-ketoacyl synthases active site"
/codon_start=1
/transl_table=1
/label="fas"
/product="putative type I fatty acid synthase"
/protein_id="CAB39571.1"
/db_xref="GI:4539127"
/translation="MTHTHDOVASDRNGNSIHGSRALADPKAGPEYVAFGSGSA
WLETFELVSSAGLEADLATIYCEVELLEVAKEIYVVRPTGEPIDWRALLAEEL
VPSDKHTSAASVPGVLLTQIAGRALARGMDLITPPVIGTGHSGVLAVALKA
GARDAPELLAAOLIGAGTIVARRKGISVIGDRPPVSNADPERIRLLDFAD
VRTVPELVIATIRNGRSVITGTPEDLSRPERKROISDKKEERKRIRGSGDFPAG
FDPVVEIGFPHLADGICIVGKAEKVGIDVTLARLTALILVRGVDMLRTVRAH
GAGVMIIDIGDGLITRLTAPVIRGIGVIVPANRGCGRTLTGCAVPPVAVAMIS
VAPTYVQIPDRIKISTKFTRLTGRSPILLKGMTPTTYDAHIVAAANAGMALIAGC
GQVEITFANREVDLSGLEPRTYQFENALGLDPLMLKLVQGGRLVOKAKOSGAID
GVVISGILDEDAVELLEELGIGISYVVRPCTTQIRSVIRIATEMSKPIVMHV

```

```

Query Match      64.0%; Score 16; DB 1; Length 43839;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      5  ttcgatctctgcaag 20
Db      31291 ttcgatctctgcaag 31276

RESULT 21
AC017911
LOCUS      AC017911 118100 bp DNA
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC017911.1 GI:6553279
VERSION    AC017911.1
KEYWORDS   HTG; HTGS_PHASE2.
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
            Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
            This sequence was identified as CDM:10212606 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            Location/Qualifiers
                source          1..118100
                               /organism="Drosophila melanogaster"
                               /db_xref="taxon:7227"
BASE COUNT  32824 a 26059 c 26550 g 32667 t
ORIGIN
Query Match      64.0%; Score 16; DB 2; Length 118100;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1  gcatlctgcatcgtcgt 16
Db      63425 GCATTGCACTCTCGT 63440

RESULT 22
AC007476/c
LOCUS      AC007476 171705 bp DNA
DEFINITION Drosophila melanogaster, chromosome 2R, region 49A-49B, BAC clone
BACR37P06, complete sequence.
ACCESSION AC007476
VERSION    AC007476.5 GI:13374640
KEYWORDS   HTG.
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
            1 (bases 1 to 171705)
            Celisner, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H.,
            Holt, R.A., Evans, C.A., Gocayne, J.D., Amanalides, P.G., Brandon, R.C.,
            Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A.,
            Carlson, J.W., Center, A., Chemp, M., Davenport, L.B., Dietz, S.M.,
            Dodson, K., Dorsett, Y., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D.,
            Ferreira, S., Fitse, E., Galie, R.F., Gay, N.S., George, R.A.,
            Gonzalez, M., Houck, J., Hoskins, R.A., Hostland, T.J.,

```

IDEHAM, C., JALALI, M., KRUSE, D., LI, P., MATTLE, B., MOSIRELI, A.,
MOTOKOSH, T. C., MOY, M., MURPHY, B., NELSON, C., NELSON, K. A., NUNOO, J.,
PACIEL, J., PARAGAS, V., PARK, S., PATEL, S., PFELTER, B.,
PHOUNEAVONG, S., PLITMAN, G. S., PURI, V., RICHARDS, S., SCHEELER, F.,
SCAPLETON, M., STRONG, R., SVETIKAS, R., TECTOR, C., WILLIAMS, S. M.,
ZAVELI, J. S., SMITH, H. O., RUBIN, G. M. and VENTER, J. C.
Sequencing of Drosophila chromosome 2R, region 49A-49B
2 (bases 1 to 171705)
Celniker, S. E., Acheyani, A., Arcaina, T. T., Baxter, E., Blazek, R. G.,
AUTHORS

/gene="Cam"
 /note="Cam gene product [alt 1]"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 /protein_id="AAF58544.1"
 /db_xref="GI:7303488"
 /translation="MADQLTEEOIAEFKAFSLFDKDGDTITTKELGTVMKSLGONP
 TEALQDMINEVDAGKLTDEEVDGMTREADIDGDSQVYEEFVTMTSK"
 join(1153..1155,5710..5884,10160..10402,14732..14760)
 /gene="Cam"
 /note="Cam gene product [alt 4]"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 /protein_id="AAF58544.1"
 /db_xref="GI:7303488"
 /translation="MADQLTEEOIAEFKAFSLFDKDGDTITTKELGTVMKSLGONP
 TEALQDMINEVDAGNGTIDPEEFLTMARKKKDTSDEELREAFRVFDKDGNGFIS
 AAELRHWTNLGKELTDEEVDGMTREADIDGDSQVYEEFVTMTSK"
 join(1153..1155,5710..5884,10160..10402,14732..14760)
 /gene="Cam"
 /note="Cam gene product [alt 2]"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 /protein_id="AAF58544.1"
 /db_xref="GI:7303488"
 /translation="MADQLTEEOIAEFKAFSLFDKDGDTITTKELGTVMKSLGONP
 TEALQDMINEVDAGNGTIDPEEFLTMARKKKDTSDEELREAFRVFDKDGNGFIS
 AAELRHWTNLGKELTDEEVDGMTREADIDGDSQVYEEFVTMTSK"
 join(5527..5642,5710..5884,10160..10402,14732..15192)
 /gene="Cam"
 /product="CT24787"
 /db_xref="FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 join(5540..5642,5710..5884,10160..10402,14732..14760)
 /gene="Cam"
 /note="Cam gene product [alt 3]"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 /protein_id="AAF58543.1"
 /db_xref="GI:7303487"
 /translation="MADQLTEEOIAEFKAFSLFDKDGDTITTKELGTVMKSLGONP
 TEALQDMINEVDAGNGTIDPEEFLTMARKKKDTSDEELREAFRVFDKDGNGFIS
 AAELRHWTNLGKELTDEEVDGMTREADIDGDSQVYEEFVTMTSK"
 join(<21254..21664,21584..>21600)
 /gene="CG13166"
 /product="CT32407"
 /db_xref="FLYBASE:FBan0013166"
 /db_xref="FLYBASE:FBgn0040758"
 /evidence=not experimental
 <21254..>21600
 /gene="CG13166"
 /map="48F5-48F6"
 /db_xref="FLYBASE:FBan0013166"
 /db_xref="FLYBASE:FBgn0040758"
 /evidence=not experimental
 /protein_id="AAF58540.1"
 /db_xref="GI:7303484"
 /translation="MSERGSAGVATGGHNAKVSPTPGSGKIAYTPCPRSEPPDIAL
 APPPKKALAIPIMSLDTANGPCTEKLHPN"
 join(<34383..34566,36219..36635,37304..37431,37529..37772
 37837..37951,38022..38163,38232..38305,38481..>38679)
 /gene="CG13165"

gene

CDS

gene

gene

mRNA

gene

Query Match
Best Local Similarity 100.0%; Score 16; DB 3; Length 266133;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaattcgatccctcgt 16
|||||
Db 158787 GCATTTCGATCCTCGT 158772

RESULT 24
LOCUS MLEPRTN5 332450 bp DNA BCT 20-FEB-2001
DEFINITION Mycobacterium leprae strain TN complete genome; segment 5/10.
ACCESSION AL583921 AL450380
VERSION AL583921.1 GI:113093133

SOURCE
ORGANISM Mycobacterium leprae.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
Mycobacterium.

REFERENCE
AUTHORS Cole,S.T., Eigmeier,K., Parkhill,J., James,K.D., Thomson,N.R., Wheeler,P.R., Honore,N., Ganley,T., Churcher,C., Harris,D., Mungall,K.K., Basham,D., Brown,D., Chillingworth,T., Connor,R., Davies,R.M., Devlin,K., Duthoy,S., Feltyell,T., Fraser,A., Hamlin,N., Holroyd,S., Hornsby,T., Jagsels,K., Lacroix,C., Maclean,J., Moule,S., Murphy,L., Oliver, Quail,M.A., Rajandream,M.-A., Rutherford,K.M., Rutter,S., Seeger,K., Simon,S., Simmonds,M., Skelton,J., Squares,R., Squares,S., Stevens,K., Taylor,K., Whitehead,S., Woodward,J.R. and Barrell,B.G.
Massive gene decay in the leprosy bacillus
Nature 409 (6823), 1007-1011 (2001)

```

MEDLINE      21128732
REFERENCE    2 (bases 1 to 332450)
AUTHORS      Parkhill,J.
TITLE        Direct Submission
JOURNAL      Submitted (20-FEB-2001) Submitted on behalf of the Mycobacterium
              leprae sequencing teams, The Sanger Centre, Wellcome Trust Genome
              Campus, Hinxton, Cambridge, CB10 1SA, UK
              Unitie de Genetique
              Molculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
              75724, Paris Cedex, France. E-mail: parkhill@sanger.ac.uk

COMMENT      Notes:
              Details of M. leprae sequencing at the Sanger Centre are available
              from http://www.sanger.ac.uk/Projects/M_leprae/ A relational
              database containing the M. leprae sequences is available from
              http://genolist.pasteur.fr/Leprona/.

FEATURES     location/Qualifiers
Source       1..332450
              /organism="Mycobacterium leprae"
              /strain="TN"
              /db_xref="taxon:1769"
gene         100..1269
              /gene="ML1153"
CDS          100..1269
              /gene="ML1153"
              /note="Possible pseudogene similar to M. tuberculosis
              paralogue RV0063 (Best blastx score 343)"
              /codon_start=1
              /pseudo
              /transl_table=11
              /product="Probable oxidoreductase (pseudogene)"
              /complement(1333..2552)
              /gene="ML1154"
              /complement(1333..2552)
              /gene="ML1154"
CDS          /note="Possible pseudogene of M. tuberculosis orthologue
              RV1318c (Best blastx score 336)"
              /codon_start=1
              /pseudo
              /transl_table=11
              /product="hypothetical protein (pseudogene)"
              2959..3618
              /gene="ML1155"
              /note="Pfam match to entry PF01939 DUF91, protein of
              unknown function, score 445.00, E-value 6.6e-130"
              2959..3621
              /gene="ML1155"
              /note="Previously sequenced Mycobacterium leprae
              hypothetical protein SW:YD21_MYCLE (P53524) (201 aa);
              FastA score E(): 0, 99.5% identity in 193 aa overlap. Also
              highly similar to Mycobacterium tuberculosis hypothetical
              protein RV1321 SW:YD21_MYCTU (Q10634) (226 aa); FastA
              score E(): 0, 88.2% identity in 220 aa overlap. Contains
              Pfam match to entry PF01939 DUF91, Protein of unknown
              function."
              /codon_start=1
              /transl_table=11
              /product="conserved hypothetical protein"
              /protein_id="CAC31536.1"
              /db_xref="GI:13093134"
              /translation="MIAQCTVDYLGRLTAHPSARILLFKADGSVVAHDBAYPL
              NMSPPCWMLTEDVAGSPAVVKNKAGEBQRTVEVEHDSCHLGVADGLKYGDEA
              HLDCTLLAEVOLLGEGYTLVRCVEMTAIGPVDLCDEGGSVAVEIKRGELDGEV
              LTRIALNLNNSVLAIPVRGVEFAOHKIPQARLTATDRGIRCVTLVDYDKMRGMSDEY
              LF"
              2959..3621
              /gene="ML1155"
              3645..3928
              /gene="ML1156"
              3645..3928
              /gene="ML1156"
              /note="Possible pseudogene of M. tuberculosis orthologue
              RV1322 (Best blastx score 222)"
              /codon_start=1
              /pseudo

misc_feature /transl_table=11
              /product="hypothetical protein (pseudogene)"
              /complement(3953..4420)
              /gene="ML1157"
              /complement(3953..4420)
              /note="Identical to the previously sequenced Mycobacterium
              leprae TR:Q49717 (EMBL:U00014) (155 aa); FastA score E():
              0, 99.4% identity in 155 aa overlap. Also similar to
              Streptomyces coelicolor conserved hypothetical protein
              TR:CA870628 (EMBL:AL137242) (146 aa); FastA score E():
              1.5e-19, 42.1% identity in 140 aa overlap and Hemophilus
              influenzae iactoyglutathione lyase (EC 4.4.1.5)
              SW:LGUL_IABEIN (P44638) (135 aa); FastA score E(): 0.69,
              27.5% identity in 138 aa overlap. Contains Pfam match to
              entry PF01626 4HPPD_C, 4-hydroxyphenylpyruvate dioxygenase
              C terminal domain."
              /codon_start=1
              /transl_table=11
              /product="conserved hypothetical protein"
              /protein_id="CAC31538.1"
              /db_xref="GI:13093135"
              /translation="MNTTDQINTCOLATTVITVTDHYGIADVADLMAISYHDHICM
              IIVHEVNDQGITREAMLSAHGAPIGTQIOIMAPINSSPIATFLDKHGCTQOLAC
              RVSDDLRLTKRRRAGCVRLVYTPFGQISMARINIFHKDAGVILIEVPAS"
              /complement(3962..4420)
              /gene="ML1157"
              /note="Pfam match to entry PF01626 4HPPD_C,
              4-hydroxyphenylpyruvate dioxygenase C terminal domain,
              score 179.90, E-value 4e-50"
              4512..5693
              /gene="fada4"
              /note="fada4"
              4512..5690
              /gene="ML1158"
              /note="fada4"
              /note="Pfam match to entry PF00108 thiolase, Thiolase,
              score 760.80, E-value 5.8e-225"
              4512..5693
              /gene="fada4"
              /note="fada4"
              /note="Identical to the previously sequenced Mycobacterium
              leprae probable acetyl-CoA acetyltransferase (EC 2.3.1.9)
              SW:THIL_MYCLE (P46707) (393 aa); FastA score E(): 0,
              100.0% identity in 393 aa overlap. Also highly similar to
              many others including: acetyl-CoA acetyltransferases from
              Alcaligenes eutrophus SW:THIL_ALCEU (P14611) (393 aa);
              FastA score E(): 0, 50.4% identity in 389 aa overlap and
              Mycobacterium tuberculosis (Fada4) RV1323 SW:THIL_MYCTU
              (Q10629) (389 aa); FastA score E(): 0, 86.9% identity in
              388 aa overlap. Contains Pfam match to entry PF00108
              thiolase, Thiolase. Contains PS00098 Thiolases acyl-enzyme
              intermediate signature. Contains PS00099 Thiolases active
              site. Contains PS00737 Thiolases signature 2.
              ML2162"
              /codon_start=1
              /transl_table=11
              /product="possible acetyl-CoA C-acetyltransferase"
              /protein_id="CAC31539.1"
              /db_xref="GI:13093136"
              /translation="MNTTSVIVTGARTPIGKLMGSLKDPFASDLGATITAAALKKANYA
              PSIVQYVILIGCVLTAGAGOMPAAQAAVAGIGMVPVLTINKMCLSLDAIATADOLI
              RAGEVDVYVAGGSGSMRKAPHLIIMDSRSGYKGYDVTIVDHLAYDGLHVDTPNMGAL
              TGEORNDVCKPTROEDDEFAARSHOKAAANKDGFADBEVYPVSIPOKSGSLDPTDEE
              GIRANTSKESLAGLPAPRCGGTITPESASISGGAATVVMNKKADQGLTWLVTI
              GAHGVVAGPDSITLOSAPANAIRKAVNDREGISVEDLDVEINERFAAVALASARLGTIA
              PELVNVNGAIAVGHPIPLMGSGARITLVALETLARSGCYVAALCGAGGOGDALIIRA
              V"
              4761..4817
              /gene="fada4"
              /note="PS00098 Thiolases acyl-enzyme intermediate
              signature"
              5526..5576
              /gene="fada4"
              /note="PS00737 Thiolases signature 2"

```

```

misc_feature 5631..5672
              /gene="fada4"
              /note="PS00099 Thiolases active site"
gene 5880..6785
      /gene="ML1159"
      /gene="ML1159"
      /gene="ML1159"
      /note="Previously sequenced Mycobacterium leprae TR:Q49716
      (EMBL:U00014) (255 aa); Fasta score E(): 0, 100.08
      identity in 217 aa overlap. Highly similar to
      Mycobacterium tuberculosis Rv1324 SW:YD24_MYCTU (Q10636)
      (304 aa); Fasta score E(): 0, 72.48 identity in 301 aa
      overlap. Also weakly similar in parts to several putative
      thiodoxins e.g. Alicyclobacillus acidocaldarius
      SW:PHO-ALICAC (P80579) (105 aa); Fasta score E(): 0.00045,
      27.18 identity in 107 aa overlap."
      /codon_start=1
      /transl_start=1
      /product="conserved hypothetical protein"
      /protein_id="CAC31540.1"
      /db_xref="GI:13093137"
      /translation="MACAVDLSGLKORAKASTSPASRATUGARCTGSENTSYIE
      ITEANFDEVLVRSHEVPLILVWSPRSQACVKLETLISGLAVASDSTWLLAVNVDA
      VPRVAQIFGVDAVPTVVALAGPPLSSPQGSVDQRLRMDSILSTAGKLRGPPRS
      EDSAEIDPAIAQARQLEAGDFLTAKOSYAILDADPASYEAKAIAIQLIDPLPATQ
      HPDAVADAPAGDIAAFAPAAADVOILNDVTAFFRLIYLIVNSTGDSERSVTRRLI
      ELPELPDPDPNVIGRRNLNANLY"
      complement(6841..9018)
      /gene="ML1160"
      /gene="ML1160"
      /gene="ML1160"
      /note="Possible pseudogene of M. tuberculosis orthologue"
Query Match 64.0%; Score 16; DB 1; Length 332450;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 ttcgactcctgcagc 20
    |||||||||||||||
Db 74557 TTCGATCCTCGTCGACG 74542

RESULT 25
LOCUS TC057884 728 bp DNA PLN 30-MAR-1997
DEFINITION Thanatephorus cucumeris strain 2R14 5.8S ribosomal RNA gene and
            internal transcribed spacers 1 and 2, complete sequence.
ACCESSION U57884
VERSION 057884.1 GI:1914814
KEYWORDS
SOURCE .
ORGANISM Thanatephorus cucumeris.
            Thanatephorus cucumeris
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes;
            Heterobasidiomycetes; Heterobasidiomycetidae; Ceratobasidiiales;
            Ceratobasidiaceae; Thanatephorus.
REFERENCE 1 (bases 1 to 728)
AUTHORS Salazar,O., Schneider,H., Keijer,J. and Rubio,V.
TITLE Phylogenetic relations in Rhizoctonia solani AG2 using ITS
sequences
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 728)
AUTHORS Rubio,V.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-1996) Victor Rubio, Biocnecologia Microbiana,
            Centro Nacional de, Biotecnologia (CSIC-UAM), Campus
            Cantoblanco-UAM, Madrid, 28049, Spain
FEATURES
source 1..728
            /organism="Thanatephorus cucumeris"
            /strain="2R14"
            /specific_host="Tulipa sp."
            /db_xref="taxon:118239"
            /tissue_type="mycelia"

```

```

misc_RNA /note="isolated from tulip, from The Netherlands"
          1..294
          /note="internal transcribed spacer 1; ITS1"
          295..449
          /product="5.8S ribosomal RNA"
          450..728
          /note="internal transcribed spacer 2; ITS2"
BASE COUNT 196 a 153 c 145 g 234 t
ORIGIN

Query Match 60.0%; Score 15; DB 8; Length 728;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 cctgcagatgacg 25
    |||||||||||||||
Db 20 CCTGCTCAGCGTCGACG 34

RESULT 26
LOCUS RICCCTALD 1801 bp DNA PLN 14-APR-2000
DEFINITION Oryza sativa gene for cytoplasmic aldolase, complete cds,
            clone:AlDC-a.
ACCESSION D13512.1 GI:218156
VERSION cytoplasmic aldolase.
KEYWORDS Oryza sativa DNA, clone:AlDC-a.
SOURCE Oryza sativa
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzaceae; Oryza.
            1 (bases 1 to 1801)
REFERENCE 1 (bases 1 to 1801)
AUTHORS Tsutsumi,K., Kagaya,Y., Hidaka,S., Suzuki,J., Tokairin,Y.,
            Hirai,T., Hu,D.L., Ishikawa,K. and Ejiri,S.
TITLE Structural analysis of the chloroplastic and cytoplasmic
            aldolase-encoding genes implicated the occurrence of multiple loci
            in rice
JOURNAL Gene 141 (2), 215-220 (1994)
MEDLINE 94215906
REFERENCE 2 (bases 1 to 1801)
AUTHORS Tsuchiya,T.
TITLE Direct Submission
JOURNAL Submitted (28-OCT-1992) to the DDBJ/EMBL/GenBank databases. Tohru
            Tsuchiya, Iwate University, Department of Agriculture, 3-18-8 Ueda,
            Morioka, Iwate 020, Japan (Tel:0196-23-5171(ex.2776),
            Fax:0196-25-7667)
COMMENT Submitted (28-OCT-1992) to DDBJ by:
            Tohru Tsuchiya
            Institute for Cell Biology and Genetics
            Faculty of Agriculture
            Iwate University
            3-18-8 Ueda
            Morioka, Iwate 020
            Japan
            Phone: 0196-23-5171 x2776
            Fax: 0196-24-5084.
FEATURES
source 1..1801
            /organism="Oryza sativa"
            /db_xref="taxon:4530"
            /clone="AlDC-a"
            41..49
            /product="cytoplasmic aldolase"
            /protein_id="BAA02729.1"
            /db_xref="GI:218157"
            /translation="MSAYCGKXKDELIKMAAYIGTPGKGLADESTGTICKRFASIN
            join(152..179,615..1663)
            /codon_start=1
            /number=1
            80..179
            /product="cytoplasmic aldolase"
            /protein_id="BAA02729.1"
            /db_xref="GI:218157"
            /translation="MSAYCGKXKDELIKMAAYIGTPGKGLADESTGTICKRFASIN

```

VENVEDNRRAFRRELLFCTPGALOYISGVILFDETLKYOKTKCKGKPFVWDILKEAGALFCI
KVDKGTIEVAGTGTETTTGGHDIGKOCATYFAGAFKRAVLKIGPQPSOLAID
LNAOGLACVAILCOENGLVPIVEPEILVDSPHIDRCAYSEVVLACAYALNHNHVL
LEGTLLKPNVTSDDAKKVAPEVIAETVTRTORVPPAVPAVPAVIFLSGSGSEETI
NLNMMKLSAKKPMWSLSFSGPALQOSTLAKMAGKIDENVEKAAAFVLRCKNSENAIL
GTYRGDAVLGEGAGESLHVADKY"

intron
180..614
/number=1
615..1801
/number=2

BASE COUNT 418 a 521 c 439 g 423 t

ORIGIN

Query Match 60.0%; Score 15; DB 8; Length 1801;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tectcgtcaggtgca 24
|||||
Db 1551 TCCTCGTCAGGTGCA 1565

RESULT 27
AB037816/c 4886 bp mRNA PRI 14-MAK-2000
LOCUS Homo sapiens mRNA for KIAA1395 protein, partial cds.
DEFINITION AB037816
VERSION AB037816.1 GI:7243170
KEYWORDS
SOURCE homo sapiens, brain cDNA to mRNA, clone_lib:phluescriptII SK plus
clone:hj07927.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Nagase,T., Kikuno,R., Ishikawa,K.I., Hirosewa,M. and Ohara,O.
TITLE 1 (sites)
JOURNAL Prediction of the coding sequences of unidentified human genes.
AUTHORS XVI. The complete sequences of 150 new cDNA clones from brain which
TITLE code for large proteins in vitro
JOURNAL DNA Res. 7 (1), 65-73 (2000)
20181126
2 (bases 1 to 4886)
Ohara,O., Nagase,T. and Kikuno,R.
REFERENCE Direct Submission
AUTHORS Submitted (31-JAN-2000) to the DDBJ/EMBL/Genbank databases. Osamu
TITLE Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology;
JOURNAL 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:cdnainfo@kazusa.or.jp, URL:http://www.kazusa.or.jp/huge/,
Tel:+81-438-52-3913, Fax:+81-438-52-3914)

FEATURES
source
1..4886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hj07927"
/clone_lib="phluescriptII SK plus"
/tissue_type="brain"
1..4884
/gene="KIAA1395"
/note="KIAA1395"
/note="Start codon is not identified."
/codon_start=1
/product="KIAA1395 protein"
/protein_id="BA02633.1"
/db_xref="GI:7243171"
/translation="DARTMAASERRAFKINRTVAAEVRKQVRSRSGSPHSRRCS
SLGVLPEVPELDFEDVLLSRPDAPGRLDVEFPADDELILLQRECRTEG
IPKDKIDAOVRAVEMYIEDMVIVHRRGYLSAASPTTDTQEROKGLPROVEPQ
DASDERSGSPDSNRRSGSSPEDTPRSSASSIFDLRLAASDLPLLERAEPD
VDRNETLRQHRPPLLTLYPAPDEDAVERGSRPPEPRHFGORILVCLSKETI
EIEPFGILAYDVERKKIISENFPDI NSDSMKGLRAQTHPAISTLAKSAIFSVI
YSPDIFLIVIKLEKVLQGDISECCPEYVWLKEVDIAKNEKELEKRLAEOPCSTRIG

RYRMPAMTAVHLANIYSSAQLDRDSSEGERPAMTDRRRRGPDRASSGDACSE
SGFRPATLITVINEFKOAEKLSDEDLFEFLADMRRPSSLRLRPPVIAQLIDISAP
ENHRLSPBELLHIKPYDDPRGPRLKTEFPARVAVPHTSYNNLXYVPHSNFSS
RGSYRNILAVROYMTGEDPQALVIEGSSCSPTREATPVYHNNKSPFEYEEK
LHLPACVTENHLLFTFYHVSQPRPGALFPPGFTWIPILQHRRLRTGFCPLVAV
DOPPSYSLVTPDVALPGMRVWDGKKGVSEILNAVSSVHNPQDYLDKFTLVVLE
GAFPRLDVTIVSEGVDELKASLALRLASPELVAFSHHVLDKLRLVTRPILIS
GQIVMIGRAPEAMAHVSVLHRSLEAQRGHCPOQLAAVVAHAFRLPGTEPSLPQ
APPVVOATLARGSGRPSALYLAKSKISSNDPLVAPSGVDEVSRIILAKLHE
FLALOMVSSSAVREAILOHAFEFOLKSMALLHLLGLDLPTRKIRFGRFLDIT
TALGVSGLEVITRKHVDLAEHNLASLAFPLDLSLVDRGFVFLVRAHYOVAT
RUOSSPNALITLMEPTRTICSHENVTINLCCCPSPASPSVSTTQSSSF
SSQADPDKTSMFEISGPPRQOHLFGLLTELALALEPBAEGAPLAKKAKISVHSL
LCGHDPDPRYATAVKARVAELVPLLSIARDTLPRLNDPBAEGRQSRSLMDSOT
BEGGDIACTINPSVAMALAGFLAPGSRASISQGPPTASRAGCALSAESSRTLLACVL
VWLKNTPEALLORNAEDTLTPOIGRLDLILVLCIAAFYKKAKAFERINSTRLLCYL
MKARLEAILGTIGAROEVRYSRSPFGNPEVBMKSVTHNKQTSDRDKTDKM
EHEALVEGNLATEASLVLDITLILVYVMSLPARESYLGAIVLVVYLSGASQALP
LOHGLATOKALVSKPELLFEDDTLCAIDLRLRLRGGSRISTTRHASSLYLAKR
ONFETGHMFAKMOVTWSSLSVGTGTONSEEHRLRSKTLITLYAEEDMLRSTEA
EOVDLMENLHMLITDITVAKMEHEDPMLIDLMYRIARGVGSPLDLTWMOMAGK
HAELGNHAEAAQCVHAA"

BASE COUNT 941 a 1605 c 1425 g 915 t

ORIGIN

Query Match 60.0%; Score 15; DB 9; Length 4886;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 atctcgtcaggtgc 23
|||||
Db 735 ATCCTCGTCAGGTGC 721

RESULT 28
AX173175/c 6372 bp DNA PAT 03-JUL-2001
LOCUS AX173175
DEFINITION Sequence 1 from Patent WO0142295.
ACCESSION AX173175
VERSION AX173175.1 GI:14598036
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 6372)
TITLE Lu,P., Garman,J.D. and Candia,A.F.
JOURNAL Clasp-7 transmembrane protein
Patent: WO 0142295-A.1 14-JUN-2001;
Atdor Vita Corporation (US)

FEATURES
source
1..6372
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="full length human CLASP-7 cDNA"
13..6156
/note="human CLASP-7"
/codon_start=1
/protein_id="CAC43731.1"
/db_xref="GI:14598037"
/translation="MAASERRAFKINRTVAAEVRKQVRSRSGSPHSRRCSLSG
VPLTEVPELDFEDVLLSRPDAPGRLDVEFPADDELILLQRECRTEPQPKD
EKIDAOVRAVEMYIEDMVIVHRRGYLSAASPTTDTQEROKGLPROVEPQDASC
DERSGSPDSNRRSGSPEDTPRSSGASITDELNLNADSLPLSLERAABEVDOR
NETLRQHRPPLLTLYPAPDEDAVERGSRPPEPRHFGORILVCLSKFETIEP
IFGILAYDVERKKIISENFPDI NSDSMKGLRAHGTHPAISTLARSATISVYPS
DIEVLVIEKLVLOOGDITSPCCPEYVWLKEVDIAKNEKELEKRLAEOFCRGLGRM
PRAMTAVHLANIYSSAQLDRDSSEGERPAMTDRRRRGPDRASSGDACSEGR
PATLITVINEFKOAEKLSDEDLFEFLADMRRPSSLRLRPPVIAQLIDISAPNH
PCLSPDLITIKYPRDPRKTELEFPAREVYAHVTSYRLITLYPSLWFSRQGS
VNLNARVOYMTGEDPQALVIEGSSCSSETPRATPPVYHNNKSPFEYEFKILP
ACVTENHLLFTFYHVSQPRPGALFPPGFTWIPILQHRGLRTGFCPLVSDQRP

of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone
RP13-379024. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.

The true right end of clone RP13-379024 is at 59999 in this
sequence. The true right end of clone RP5-908M14 is at 100 in this
sequence.

FEATURES

source
1. .59999
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="RP13-379024"
/clone_1b="RRC1-13.2"
complement(1..95)
/note="match: STS: Em:HS908M14"
1020. .1119
/note="2 copies 50 mer 91% conserved"
1026. .1121
/note="4 copies 24 mer 79% conserved"
3539. .3867
/note="AluJo repeat: matches 3. .306 of consensus"
3939. .4116
/note="AluJo repeat: matches 54. .220 of consensus"
5954. .6007
/note="2 copies 27 mer 96% conserved"
6334. .6456
/note="MER5A repeat: matches 1. .125 of consensus"
6931. .7062
/note="4 copies 33 mer 84% conserved"
complement(7115)
/gene="DB379024.1"
complement(join(7115..8609,8958..9082,9452..9539,
10044..10169,17021..17196,18617..>19251))
/gene="bB379024.1"
/note="match: cDNAs: Em:U84725
match: ESTs: Em:H80372 Em:H80385 Em:W00677 Em:AI615257
Em:M72525 Em:H80289 Em:AA127875 Em:AA127913 Em:AW770386
Em:AA972923 Em:AI685293 Em:AA534131 Em:AW973240
Em:AA533669 Em:AA127923 Em:N53566 Em:AW511251 Em:AI174441"
/evidence=not_experimental
/product="bB379024.1 (novel protein similar to
transcription factor GATA-5)"
complement(7115..19251)
/gene="bB379024.1"
complement(7128..7133)
/gene="bB379024.1"
7724. .8104
/note="match: STS: Em:G25060"
7724. .8137
/note="match: STS: Em:G29730"
7764. .7900
/note="match: STS: matches 1. .124 of consensus"
complement(join(8454..8609,8958..9082,9452..9539,
10044..10169,17021..17196,18617..19139))
/gene="bB379024.1"
/note="match: proteins: Tr:Q16365 Sw:P43692 Wp:F52C12
Tr:O9TXV3 Sw:P19212 Sw:P23772 Sw:P23825 Sw:O91428
Tr:O9Y7E8 Sw:O13415 Tr:O77157 Sw:P28515 Sw:P43695
Sw:O91678 Sw:O92908 Sw:O91677 Sw:P70005 Sw:P23773
Sw:P23770 Sw:P17679 Tr:O90410 Sw:P17678 Sw:P43429"
/codon_start=1
/evidence=not_experimental
/product="bB379024.1 (novel protein similar to
transcription factor GATA-5)"
protein_id="CAC36001.1"
/db_xref="GI:13559069"
/translation="MYOSLALASPRQAAAYADSGSLHAPQAGSPMFVPPAPVPSMUS
YLSCEPPEPPELAPRCWAQATTAASAGPSPPIPPAHPPAGATAPFAISPSGP
GSGGAGRDSDAVOAGLIPREDFAPALGRVGYTSATVPAYVSPVADSWTAPPD
GSVHLGLPGRPRFTVSDPLFEFPGEGRVCVNCALSTPLMRDCTGHYLCNACGLYHK

misc_feature
9836. .10334
/note="CpG island"
/evidence=not_experimental
10311. .10434
/note="MIR repeat: matches 62. .188 of consensus"
11461. .11560
/note="2 copies 50 mer 86% conserved"
11596. .11651
/note="2 copies 28 mer 94% conserved"
11672. .11727
/note="2 copies 28 mer 94% conserved"
11744. .11799
/note="2 copies 28 mer 94% conserved"
11820. .11875
/note="2 copies 28 mer 96% conserved"
11892. .11947
/note="2 copies 28 mer 96% conserved"
12774. .12958
/note="MER53 repeat: matches 10. .189 of consensus"
14602. .14773
/note="MIR repeat: matches 37. .214 of consensus"
15051. .15342
/note="AluJb repeat: matches 7. .299 of consensus"
15351. .15663
/note="AluJo repeat: matches 1. .296 of consensus"
17210. .17291
/note="2 copies 41 mer 90% conserved"
17861. .20459
/note="CpG island"
/evidence=not_experimental
19352. .20175
/note="Single clone region. Sequence from reads from a
short insert library derived from a single pUC clone.
Restriction digest data confirm the assembly."
19448. .19645
/note="match: GSS: Em:AQ938920"
19641. .19712
/note="3 copies 24 mer 84% conserved"
19642. .19713
/note="36 copies 2 mer cc 70% conserved"
complement(20394..20581)
/note="match: STS: Em:L30262"
20489. .20956
/note="117 copies 4 mer ttcc 66% conserved"
20492. .20722
/note="7 copies 33 mer 67% conserved"
20496. .20959
/note="29 copies 16 mer 65% conserved"
20500. .20950
/note="11 copies 41 mer 62% conserved"
20504. .20951
/note="16 copies 28 mer 66% conserved"
20724. .20789
/note="2 copies 33 mer 93% conserved"
20827. .20925
/note="3 copies 33 mer 76% conserved"
20831. .20920
/note="15 copies 6 mer ttttc 67% conserved"
20933
/note="Tandem repeat. Forced join. Gap size estimated to
be approximately 110bp by restriction digest data."
20958. .21234
/note="AluXs repeat: matches 9. .294 of consensus"
22223. .22386
/note="MIR repeat: matches 52. .226 of consensus"
22934. .23051
/note="L2 repeat: matches 2349. .2473 of consensus"
23253. .23314
/note="MIR repeat: matches 122. .183 of consensus"

```

repeat_region      23676..23871
                    /note="7 copies 28 mer 62% conserved"
repeat_region      23687..23851
                    /note="5 copies 33 mer 84% conserved"
repeat_region      23689..23877
                    /note="7 copies 27 mer 64% conserved"
repeat_region      23697..23846
                    /note="3 copies 50 mer 74% conserved"
repeat_region      23741..23868
                    /note="8 copies 16 mer 67% conserved"
repeat_region      25184..25344
                    /note="MIR repeat: matches 21..190 of consensus"
repeat_region      27265..27342

Query Match      60.0%; Score 15; DB 9; Length 59999;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9      atccctgcaagtgca 23
          |||||
Db      12634 ATCCTGTCAGTGTC 12648

RESULT 31
AP003452      168972 bp      DNA      HTG      29-MAR-2001
LOCUS      Oryza sativa chromosome 1 clone P0478H03, *** SEQUENCING IN
DEFINITION      PROGRESS ***, in ordered pieces.
ACCESSION      AP003452
VERSION      AP003452.1 GI:13486921
KEYWORDS      HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:P0478H03.
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
                Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE      Sasaki,T., Matsumoto,T. and Yamamoto,K.
                Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
                clone:P0478H03
                Published Only in Database (2001) In press
                2 (bases 1 to 168972)
AUTHORS      Sasaki,T., Matsumoto,T. and Yamamoto,K.
JOURNAL      Direct Submission
REFERENCE      Submitted (28-MAR-2001) Takuji Sasaki, National Institute of
                Agrobiological Resources, Rice Genome Research Program; Kannondai
                2-1-2, Tsukuba, Ibaraki 305-8602, Japan
                (E-mail:tsasaki@agr.affrc.go.jp, URL:http://rpg.dna.affrc.go.jp/,
                Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT      NOTE: It currently consists of 1 contigs. Gaps between the contigs
                are represented as runs of N. The order of the pieces is believed
                to be correct as given, however the sizes of the gaps between them
                are based on estimates that have provided by the submitter. This
                sequence will be replaced by the finished sequence as soon as it is
                available and the accession number will be preserved.
                * NOTE: This is a 'working draft' sequence.
                * This sequence will be replaced
                * by the finished sequence as soon as it is available and
                * the accession number will be preserved.

FEATURES
    source
        1..168972
            /organism="Oryza sativa"
            /cultivar="Nipponbare"
            /db_xref="taxon:4530"
            /chromosome="1"
            /clone="P0478H03"

BASE COUNT      46333 a 37251 c 37299 g 47789 t      300 others

ORIGIN
Query Match      60.0%; Score 15; DB 2; Length 168972;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      10      tccctgcaagtgca 24
          |||||
Db      165052 TCCTCGTCAGTGCA 165066

RESULT 32
AP003455      193530 bp      DNA      HTG      29-MAR-2001
LOCUS      Oryza sativa chromosome 1 clone P0519D04, *** SEQUENCING IN
DEFINITION      PROGRESS ***, in ordered pieces.
ACCESSION      AP003455
VERSION      AP003455.1 GI:13486924
KEYWORDS      HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:P0519D04.
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
                Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE      Sasaki,T., Matsumoto,T. and Yamamoto,K.
                Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
                clone:P0519D04
                Published Only in Database (2001) In press
                2 (bases 1 to 193530)
AUTHORS      Sasaki,T., Matsumoto,T. and Yamamoto,K.
JOURNAL      Direct Submission
REFERENCE      Submitted (28-MAR-2001) Takuji Sasaki, National Institute of
                Agrobiological Resources, Rice Genome Research Program; Kannondai
                2-1-2, Tsukuba, Ibaraki 305-8602, Japan
                (E-mail:tsasaki@agr.affrc.go.jp, URL:http://rpg.dna.affrc.go.jp/,
                Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT      NOTE: It currently consists of 1 contigs. Gaps between the contigs
                are represented as runs of N. The order of the pieces is believed
                to be correct as given, however the sizes of the gaps between them
                are based on estimates that have provided by the submitter. This
                sequence will be replaced by the finished sequence as soon as it is
                available and the accession number will be preserved.
                * NOTE: This is a 'working draft' sequence.
                * This sequence will be replaced
                * by the finished sequence as soon as it is available and
                * the accession number will be preserved.

FEATURES
    source
        1..193530
            /organism="Oryza sativa"
            /cultivar="Nipponbare"
            /db_xref="taxon:4530"
            /chromosome="1"
            /clone="P0519D04"

BASE COUNT      54949 a 41978 c 42335 g 54218 t      50 others

ORIGIN
Query Match      60.0%; Score 15; DB 2; Length 193530;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      10      tccctgcaagtgca 24
          |||||
Db      148848 TCCTCGTCAGTGCA 148834

RESULT 33
AR118251      66 bp      DNA      PAT      16-MAY-2001
LOCUS      AR118251
DEFINITION      Sequence 96 from patent US 6140490.
ACCESSION      AR118251
VERSION      AR118251.1 GI:14099157
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unknown.
REFERENCE      Unclassified.
                1 (bases 1 to 66)

```

AUTHORS Biesecker, G. and Gold, L.
 TITLE High affinity nucleic acid ligands of complement system proteins
 JOURNAL Patent: US 6140490-A 96 31-OCT-2000;
 FEATURES Location/Qualifiers
 source 1..66
 /organism="unknown"
 BASE COUNT 11 a 19 c 23 g 13 t
 ORIGIN

Query Match 56.0%; Score 14; DB 6; Length 66;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcattcctcgtcag 19
 Db 40 tgcattcctcgtcag 53

RESULT 34
 AX174645/c 211 bp DNA PAT 04-JUL-2001
 LOCUS AX174645 Sequence 77 from Patent WO0142296.
 ACCESSION AX174645
 VERSION AX174645.1 GI:14598220
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 211)
 AUTHORS Lu, P., Garman, J.D. and Candia, A.F.
 TITLE Clasp-5 transmembrane protein
 JOURNAL Patent: WO 0142296-A 77 14-JUN-2001;
 FEATURES Location/Qualifiers
 source 1..211
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="14th exon"
 BASE COUNT 34 a 59 c 68 g 50 t
 ORIGIN

Query Match 56.0%; Score 14; DB 6; Length 211;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 atccctcgcagtg 22
 Db 100 atccctcgcagtg 87

RESULT 35
 AF140181/c 483 bp DNA PLN 12-FEB-2001
 LOCUS AF140181 Filobasidiella neoformans isolate CN19.95 diphenol oxidase gene,
 DEFINITION partial cds.
 ACCESSION AF140181
 VERSION AF140181.1 GI:5734019
 KEYWORDS
 SOURCE Filobasidiella neoformans.
 ORGANISM Filobasidiella neoformans.
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;
 Tremellales; Tremellaceae; Filobasidiella.

REFERENCE 1 (bases 1 to 483)
 AUTHORS Xu, J., Vilgalys, R. and Mitchell, T.G.
 TITLE Multiple gene genealogies reveal recent dispersion and
 hybridization in the human pathogenic fungus *Cryptococcus*
 neoformans
 JOURNAL Mol. Ecol. 9 (10), 1471-1481 (2000)
 MEDLINE 20504730
 PUBMED 11050543

REFERENCE 2 (bases 1 to 483)
 AUTHORS Xu, J., Vilgalys, R.J. and Mitchell, T.G.
 TITLE Direct Submission
 JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke
 University, Research Dr., Durham, NC 27710, USA
 FEATURES Location/Qualifiers
 source 1..483
 /organism="Filobasidiella neoformans"
 /isolate="CN19.95"
 /db_xref="taxon:5207"
 join(<75..134,187..302,355..>442)
 /product="diphenol oxidase"
 join(<75..134,187..302,355..442)
 /note="laccase precursor"
 /codon_start=1
 /product="diphenol oxidase"
 /protein_id="AAD49841.1"
 /db_xref="GI:5734053"
 /translation="MHGMROKNTPYMDGVPICITOCPIPPGSGSYTYNFTISDQSGTYWM
 HSHYSNAMAADGLMGPILIVHSVHEPIORGRDYDEDRIVFTDMW"

BASE COUNT 124 a 108 c 108 g 143 t
 ORIGIN

Query Match 56.0%; Score 14; DB 8; Length 483;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 tgcattcctcgtca 18
 Db 416 tgcattcctcgtca 403

RESULT 36
 AF140179/c 509 bp DNA PLN 12-FEB-2001
 LOCUS AF140179 Filobasidiella neoformans isolate CN-C diphenol oxidase gene,
 DEFINITION partial cds.
 ACCESSION AF140179
 VERSION AF140179.1 GI:5734017
 KEYWORDS
 SOURCE Filobasidiella neoformans.
 ORGANISM Filobasidiella neoformans.
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;
 Tremellales; Tremellaceae; Filobasidiella.

REFERENCE 1 (bases 1 to 509)
 AUTHORS Xu, J., Vilgalys, R. and Mitchell, T.G.
 TITLE Multiple gene genealogies reveal recent dispersion and
 hybridization in the human pathogenic fungus *Cryptococcus*
 neoformans
 JOURNAL Mol. Ecol. 9 (10), 1471-1481 (2000)
 MEDLINE 20504730
 PUBMED 11050543

REFERENCE 2 (bases 1 to 509)
 AUTHORS Xu, J., Vilgalys, R.J. and Mitchell, T.G.
 TITLE Direct Submission
 JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke
 University, Research Dr., Durham, NC 27710, USA
 FEATURES Location/Qualifiers
 source 1..509
 /organism="Filobasidiella neoformans"
 /isolate="CN-C"
 /db_xref="taxon:5207"
 join(<86..145,198..313,366..449,503..>509)
 /product="diphenol oxidase"
 join(<86..145,198..313,366..449,503..>509)
 /note="laccase precursor"
 /codon_start=1
 /product="diphenol oxidase"
 /protein_id="AAD49839.1"
 /db_xref="GI:5734051"
 /translation="MHGMROKNTPYMDGVPICITOCPIPPGSGSYTYNFTISDQSGTYWM
 HSHYSNAMAADGLMGPILIVHSVHEPIORGRDYDEDRIVFTDMW"

BASE COUNT 135 a 108 c 121 g 145 t
 ORIGIN

Query Match 56.0%; Score 14; DB 8; Length 509;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ttgcattcctgca 18
 |||
 DB 427 TTGCATCCTGCTCA 414

RESULT 37
 AF140176 513 bp DNA PLN 12-FEB-2001
 LOCUS AF140176/C
 DEFINITION Filobasidiella neoformans isolate Z6284 diphenol oxidase gene,
 partial cds.
 ACCESSION AF140176
 VERSION AF140176.1 GI:5734014
 KEYWORDS
 SOURCE
 ORGANISM
 Filobasidiella neoformans.
 Filobasidiella neoformans.
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;
 Tremellales; Tremellaceae; Filobasidiella.
 REFERENCE
 AUTHORS Xu,J., Vilgalys,R. and Mitchell,T.G.
 TITLE 1 (bases 1 to 513)
 Multiple gene genealogies reveal recent dispersion and
 hybridization in the human pathogenic fungus *Cryptococcus*
 neoformans
 JOURNAL Mol. Ecol. 9 (10), 1471-1481 (2000)
 MEDLINE 20504730
 PUBMED 11050543
 REFERENCE 2 (bases 1 to 513)
 AUTHORS Xu,J., Vilgalys,R.J. and Mitchell,T.G.
 TITLE Direct Submission
 JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke
 University, Research Dr., Durham, NC 27710, USA
 FEATURES
 source
 1. .513
 /organism="Filobasidiella neoformans"
 /isolate="Z6284"
 /db_xref="taxon:5207"
 join(<76..135,188..303,356..439,493..>513)
 /product="diphenol oxidase"
 join(<76..135,188..303,356..439,493..>513)
 /note="laccase precursor"
 /codon_start=1
 /product="diphenol oxidase"
 /protein_id="AAD49836.1"
 /db_xref="GI:5734048"
 /translation="MHGMROKNSPYMDGIPGIGITQCPDPGSGSYTNTSDSGSTYWW
 HSHYSNADMGELWGLIVSHVEPIORGRDYEDRIVFVSMDMHNSET"

BASE COUNT 136 a 115 c 116 g 146 t
 ORIGIN

Query Match 56.0%; Score 14; DB 8; Length 513;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ttgcattcctgca 18
 |||
 DB 417 TTGCATCCTGCTCA 404

RESULT 38
 AF140173 526 bp DNA PLN 12-FEB-2001
 LOCUS AF140173/C
 DEFINITION Filobasidiella neoformans isolate MMRL750 diphenol oxidase gene,
 partial cds.
 ACCESSION AF140173
 VERSION AF140173.1 GI:5734011
 KEYWORDS
 SOURCE

KEYWORDS
 SOURCE
 ORGANISM
 Filobasidiella neoformans.
 Filobasidiella neoformans.
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;
 Tremellales; Tremellaceae; Filobasidiella.
 REFERENCE
 AUTHORS Xu,J., Vilgalys,R. and Mitchell,T.G.
 TITLE 1 (bases 1 to 526)
 Multiple gene genealogies reveal recent dispersion and
 hybridization in the human pathogenic fungus *Cryptococcus*
 neoformans
 JOURNAL Mol. Ecol. 9 (10), 1471-1481 (2000)
 MEDLINE 20504730
 PUBMED 11050543
 REFERENCE 2 (bases 1 to 526)
 AUTHORS Xu,J., Vilgalys,R.J. and Mitchell,T.G.
 TITLE Direct Submission
 JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke
 University, Research Dr., Durham, NC 27710, USA
 FEATURES
 source
 1. .526
 /organism="Filobasidiella neoformans"
 /isolate="MMRL750"
 /db_xref="taxon:5207"
 join(<96..155,206..321,376..459,512..>526)
 /product="diphenol oxidase"
 join(<96..155,206..321,376..459,512..>526)
 /note="laccase precursor"
 /codon_start=1
 /product="diphenol oxidase"
 /protein_id="AAD49833.1"
 /db_xref="GI:5734045"
 /translation="MHGRLQCTAFMDSVPSITQCPDPGSGSYTNTSDSGSTYWW
 HSHYSNADMGELWGLIVSHVEPIORGRDYEDRIVFVSMDMHNSET"

BASE COUNT 144 a 121 c 116 g 145 t
 ORIGIN

Query Match 56.0%; Score 14; DB 8; Length 526;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ttgcattcctgca 18
 |||
 DB 437 TTGCATCCTGCTCA 424

RESULT 39
 AF140183 528 bp DNA PLN 12-FEB-2001
 LOCUS AF140183/C
 DEFINITION Filobasidiella neoformans isolate CN140.97 diphenol oxidase gene,
 partial cds.
 ACCESSION AF140183
 VERSION AF140183.1 GI:5734021
 KEYWORDS
 SOURCE
 ORGANISM
 Filobasidiella neoformans.
 Filobasidiella neoformans.
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;
 Tremellales; Tremellaceae; Filobasidiella.
 REFERENCE
 AUTHORS Xu,J., Vilgalys,R. and Mitchell,T.G.
 TITLE 1 (bases 1 to 528)
 Multiple gene genealogies reveal recent dispersion and
 hybridization in the human pathogenic fungus *Cryptococcus*
 neoformans
 JOURNAL Mol. Ecol. 9 (10), 1471-1481 (2000)
 MEDLINE 20504730
 PUBMED 11050543
 REFERENCE 2 (bases 1 to 528)
 AUTHORS Xu,J., Vilgalys,R.J. and Mitchell,T.G.
 TITLE Direct Submission
 JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke
 University, Research Dr., Durham, NC 27710, USA
 FEATURES
 source
 1. .528

/organism="Filobasidiella neoformans"
/isolate="CN140.97"
/db_xref="taxon:5207"
join(<96, .155,208, .323,376, .459,513, .>528)
/product="diphenol oxidase"
join(<96, .155,208, .323,376, .459,513, .>528)
/note="laccase precursor"
/codon_start=1
/product="diphenol oxidase"
/protein_id="AAD49843.1"
/db_xref="GI:5734055"
/translation="MHGMROKNTPYMDVPGITQCP1PPGSSFTYQFTVSHSGTFVW
HSHYSNMAAGIWCPLIHSPEPLORGRDVEDRIVFTDWMHDDSE"

BASE COUNT 140 a 115 c 123 g 150 t
ORIGIN

Query Match 56.0%; Score 14; DB 8; Length 528;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ttgcagctcgtca 18
|||||
Db 437 TTGCATCCTCGTCA 424

RESULT 40
AF140164 529 bp DNA PLN 12-FEB-2001
LOCUS Filobasidiella neoformans isolate CN124.91 diphenol oxidase gene,
partial cds.
ACCESSION AF140164
VERSION AF140164
KEYWORDS
SOURCE
ORGANISM
Filobasidiella neoformans.
Filobasidiella neoformans
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;
Tremellales; Tremellaceae; Filobasidiella.
CDS
1 (bases 1 to 529)
Xu,J., Vilgalys,R. and Mitchell,T.G.
Multiple gene genealogies reveal recent dispersion and
hybridization in the human pathogenic fungus Cryptococcus
neoformans
JOURNAL Mol. Ecol. 9 (10), 1471-1481 (2000)
MEDLINE 20504730
PUBMED 11050543
REFERENCE 2 (bases 1 to 529)
AUTHORS Xu,J., Vilgalys,R. and Mitchell,T.G.
TITLE Direct Submission
JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke
University, Research Dr., Durham, NC 27710, USA
FEATURES
source
1..529
/organism="Filobasidiella neoformans"
/isolate="CN124.91"
/db_xref="taxon:5207"
join(<93, .152,203, .318,373, .456,509, .>529)
/product="diphenol oxidase"
join(<93, .152,203, .318,373, .456,509, .>529)
/note="laccase precursor"
/codon_start=1
/product="diphenol oxidase"
/protein_id="AAD49824.1"
/db_xref="GI:5734036"
/translation="MHGRLGTAFMDVPGITQCP1PPGSSFTYQFTVSHSGTFVW
HSHYSNMAAGIWCPLIHSPEPLORGRDVEDRIVFTDWMHDDSEV"

BASE COUNT 144 a 121 c 118 g 146 t
ORIGIN

Query Match 56.0%; Score 14; DB 8; Length 529;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ttgcagctcgtca 18
|||||
Db 434 TTGCATCCTCGTCA 421

RESULT 41
AF140165 530 bp DNA PLN 12-FEB-2001
LOCUS Filobasidiella neoformans isolate CN2109.91 diphenol oxidase gene,
partial cds.
ACCESSION AF140165
VERSION AF140165
KEYWORDS
SOURCE
ORGANISM
Filobasidiella neoformans.
Filobasidiella neoformans
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;
Tremellales; Tremellaceae; Filobasidiella.
CDS
1 (bases 1 to 530)
Xu,J., Vilgalys,R. and Mitchell,T.G.
Multiple gene genealogies reveal recent dispersion and
hybridization in the human pathogenic fungus Cryptococcus
neoformans
JOURNAL Mol. Ecol. 9 (10), 1471-1481 (2000)
MEDLINE 20504730
PUBMED 11050543
REFERENCE 2 (bases 1 to 530)
AUTHORS Xu,J., Vilgalys,R. and Mitchell,T.G.
TITLE Direct Submission
JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke
University, Research Dr., Durham, NC 27710, USA
FEATURES
source
1..530
/organism="Filobasidiella neoformans"
/isolate="CN2109.91"
/db_xref="taxon:5207"
join(<94, .153,204, .319,374, .457,510, .>530)
/product="diphenol oxidase"
join(<94, .153,204, .319,374, .457,510, .>530)
/note="laccase precursor"
/codon_start=1
/product="diphenol oxidase"
/protein_id="AAD49825.1"
/db_xref="GI:5734037"
/translation="MHGRLGTAFMDVPGITQCP1PPGSSFTYQFTVSHSGTFVW
HSHYSNMAAGIWCPLIHSPEPLORGRDVEDRIVFTDWMHDDSEV"

BASE COUNT 144 a 122 c 118 g 146 t
ORIGIN

Query Match 56.0%; Score 14; DB 8; Length 530;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ttgcagctcgtca 18
|||||
Db 435 TTGCATCCTCGTCA 422

RESULT 42
AF140169 530 bp DNA PLN 12-FEB-2001
LOCUS Filobasidiella neoformans isolate J10 diphenol oxidase gene,
partial cds.
ACCESSION AF140169
VERSION AF140169
KEYWORDS
SOURCE
ORGANISM
Filobasidiella neoformans.
Filobasidiella neoformans
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;
Tremellales; Tremellaceae; Filobasidiella.
CDS
1 (bases 1 to 530)
Xu,J., Vilgalys,R. and Mitchell,T.G.

TITLE Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans*

JOURNAL Mol. Ecol. 9 (10), 1471-1481 (2000)

MEDLINE 20504730

PUBMED 11050543

REFERENCE 2 (bases 1 to 530)

AUTHORS Xu,J., Vilgalys,R.J. and Mitchell,T.G.

TITLE Direct Submission

JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke University, Research Dr., Durham, NC 27710, USA

FEATURES

source

1..530

/organism="Filobasidiella neoformans"

/isolate="J10"

/db_xref="taxon:5207"

/join(<94..153,204..319,374..457,510..>530)

/product="diphenol oxidase"

/join(<94..153,204..319,374..457,510..>530)

/note="laccase precursor"

/codon_start=1

/product="diphenol oxidase"

/protein_id="AAD49829.1"

/db_xref="GI:5734041"

/translation="WHGLROLGTAFMDVPGITQCP1PPGSSFTYQFTVSHSGTFMW HSHVNSMADGIMGPLIHSPNEPLGRGRDYDEDRIVFTDWVHNSSEV"

BASE COUNT 144 a 122 c 119 g 145 t

ORIGIN

Query Match 56.0%; Score 14; DB 8; Length 530;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ttgcactcgtca 18
|||||

Db 435 TTGCATCCTCGTCA 422

RESULT 43

AF140172/c 530 bp DNA PLN 12-FEB-2001

LOCUS Filobasidiella neoformans isolate M0061 diphenol oxidase gene, partial cds.

DEFINITION AF140172

ACCESSION AF140172.1 GI:5734010

KEYWORDS

SOURCE Filobasidiella neoformans.

ORGANISM Filobasidiella neoformans

REFERENCE Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.

AUTHORS Xu,J., Vilgalys,R. and Mitchell,T.G.

TITLE 1 (bases 1 to 530)

JOURNAL Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans*

MOL. ECOL. 9 (10), 1471-1481 (2000)

MEDLINE 20504730

PUBMED 11050543

REFERENCE 2 (bases 1 to 530)

AUTHORS Xu,J., Vilgalys,R.J. and Mitchell,T.G.

TITLE Direct Submission

JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke University, Research Dr., Durham, NC 27710, USA

FEATURES

source

1..530

/organism="Filobasidiella neoformans"

/isolate="M0061"

/db_xref="taxon:5207"

/join(<94..153,204..319,374..457,510..>530)

/product="diphenol oxidase"

/join(<94..153,204..319,374..457,510..>530)

/note="laccase precursor"

/codon_start=1

/product="diphenol oxidase"

/protein_id="AAD49835.1"

/db_xref="GI:5734047"

/translation="WHGLROLGTAFMDVPGITQCP1PPGSSFTYQFTVSHSGTFMW HSHVNSMADGIMGPLIHSPNEPLGRGRDYDEDRIVFTDWVHNSSEV"

BASE COUNT 144 a 122 c 118 g 146 t

ORIGIN

/codon_start=1

/product="diphenol oxidase"

/protein_id="AAD49832.1"

/db_xref="GI:5734044"

/translation="WHGLROLGTAFMDVPGITQCP1PPGSSFTYQFTVSHSGTFMW HSHVNSMADGIMGPLIHSPNEPLGRGRDYDEDRIVFTDWVHNSSEV"

BASE COUNT 144 a 123 c 118 g 145 t

ORIGIN

Query Match 56.0%; Score 14; DB 8; Length 530;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ttgcactcgtca 18
|||||

Db 435 TTGCATCCTCGTCA 422

RESULT 44

AF140175 530 bp DNA PLN 12-FEB-2001

LOCUS Filobasidiella neoformans isolate ZG280 diphenol oxidase gene, partial cds.

DEFINITION AF140175

ACCESSION AF140175.1 GI:5734013

KEYWORDS

SOURCE Filobasidiella neoformans.

ORGANISM Filobasidiella neoformans

REFERENCE Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.

AUTHORS Xu,J., Vilgalys,R. and Mitchell,T.G.

TITLE 1 (bases 1 to 530)

JOURNAL Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans*

MOL. ECOL. 9 (10), 1471-1481 (2000)

MEDLINE 20504730

PUBMED 11050543

REFERENCE 2 (bases 1 to 530)

AUTHORS Xu,J., Vilgalys,R.J. and Mitchell,T.G.

TITLE Direct Submission

JOURNAL Submitted (01-APR-1999) Department of Microbiology, Duke University, Research Dr., Durham, NC 27710, USA

FEATURES

source

1..530

/organism="Filobasidiella neoformans"

/isolate="ZG280"

/db_xref="taxon:5207"

/join(<94..153,204..319,374..457,510..>530)

/product="diphenol oxidase"

/join(<94..153,204..319,374..457,510..>530)

/note="laccase precursor"

/codon_start=1

/product="diphenol oxidase"

/protein_id="AAD49835.1"

/db_xref="GI:5734047"

/translation="WHGLROLGTAFMDVPGITQCP1PPGSSFTYQFTVSHSGTFMW HSHVNSMADGIMGPLIHSPNEPLGRGRDYDEDRIVFTDWVHNSSEV"

BASE COUNT 144 a 122 c 118 g 146 t

ORIGIN

Query Match 56.0%; Score 14; DB 8; Length 530;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ttgcactcgtca 18
|||||

Db 435 TTGCATCCTCGTCA 422

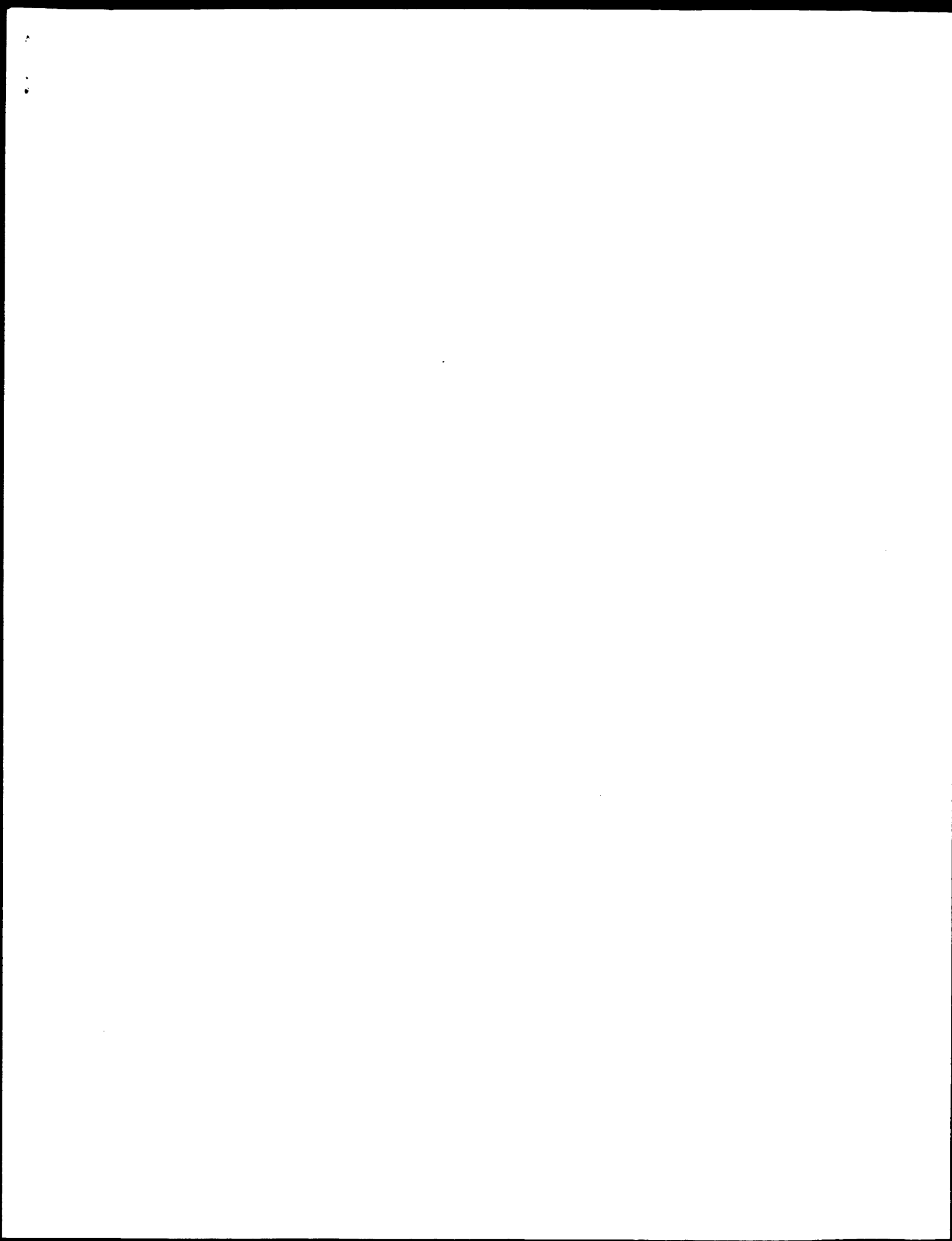
RESULT 45


```

AF140160/c  AF140160      531 bp      DNA      PLN      12-FEB-2001
LOCUS
DEFINITION  Filobasidiella neoformans isolate CN196.98 diphenol oxidase gene,
partial cds.
ACCESSION  AF140160
VERSION
KEYWORDS
SOURCE      AF140160.1  GI:5733998
ORGANISM    Filobasidiella neoformans.
            Filobasidiella neoformans.
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellomycetidae;
            Tremellales; Tremellaceae; Filobasidiella.
REFERENCE   1  (bases 1 to 531)
AUTHORS     Xu,J., Vilgalys,R. and Mitchell,T.G.
TITLE       Multiple gene genealogies reveal recent dispersion and
            hybridization in the human pathogenic fungus Cryptococcus
            neoformans
JOURNAL     MOL. ECOL. 9 (10), 1471-1481 (2000)
MEDLINE    20504730
PUBMED     11050543
REFERENCE   2  (bases 1 to 531)
AUTHORS     Xu,J., Vilgalys,R.J. and Mitchell,T.G.
TITLE       Direct Submission
JOURNAL     Submitted (01-APR-1999) Department of Microbiology, Duke
            University, Research Dr., Durham, NC 27710, USA
FEATURES
            location/Qualifiers
            1..531
            /organism="Filobasidiella neoformans"
            /isolate="CN196.98"
            /db_xref="taxon:5207"
            join(<95..154,205..320,375..458,511..>531)
            /product="diphenol oxidase"
            join(<95..154,205..320,375..458,511..>531)
            /note="laccase precursor"
            /codon_start=1
            /product="diphenol oxidase"
            /protein_id="AAD49820.1"
            /db_xref="GI:5734032"
            /translation="MHGIRQLGTAFMDGVPGITQCPIPPGSSFTYQFTVSHQSGTFMW
            HSHYSNSMADQIWGPIIHSFNPLOGRGRDYEDRIVFTIDWVHDSNV"
CDS
            mRNA
            join(<95..154,205..320,375..458,511..>531)
            /product="diphenol oxidase"
            join(<95..154,205..320,375..458,511..>531)
            /note="laccase precursor"
            /codon_start=1
            /product="diphenol oxidase"
            /protein_id="AAD49820.1"
            /db_xref="GI:5734032"
            /translation="MHGIRQLGTAFMDGVPGITQCPIPPGSSFTYQFTVSHQSGTFMW
            HSHYSNSMADQIWGPIIHSFNPLOGRGRDYEDRIVFTIDWVHDSNV"
BASE COUNT  144 a      122 c      118 g      147 t
ORIGIN

```

Search completed: December 26, 2001, 12:53:07
Job time: 8909 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:16:57 : Search time 368.09 seconds
(without alignments)
58.228 Million cell updates/sec

Title: US-09-396-196f-8

Perfect score: 25

Sequence: 1 gcatctgcgtcctcgtcagatcag 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size: 9

Total number of hits satisfying chosen parameters: 31763

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : N_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	839	22	AA162941 Human genomic DNA
2	25	100.0	1041	20	AA101303 E. coli bio B gene
3	25	100.0	1084	10	AA191329 E. coli bio B gene
4	25	100.0	1121	7	AA160496 Biotin-biosynthesis
5	25	100.0	5872	15	AA062386 Sense primer to am
6	15	60.0	34	16	AA075832 Human OREF ORF1118
7	15	60.0	586	21	AA075832 Human OREF ORF1118
8	15	60.0	6372	22	AA155684 Human CLASP-7 enco
9	15	60.0	66	20	AA219643 Complement system
10	15	60.0	370	22	AA134737 Human colon cancer
11	15	60.0	974	21	AA134737 Arabidopsis thalia

12	14	56.0	974	21	AA162941	Human genomic DNA
13	14	56.0	1160	22	AA101303	E. coli bio B gene
14	14	56.0	1302	22	AA191329	E. coli bio B gene
15	14	56.0	1425	22	AA062386	Sense primer to am
16	14	56.0	1758	19	AA075832	Human OREF ORF1118
17	14	56.0	2014	22	AA075832	Human OREF ORF1118
18	14	56.0	2104	17	AA075832	Human OREF ORF1118
19	14	56.0	2866	22	AA155684	Human CLASP-7 enco
20	14	56.0	3720	21	AA219643	Complement system
21	14	56.0	4027	22	AA134737	Human colon cancer
22	14	56.0	7215	22	AA134737	Human colon cancer
23	14	56.0	66685	22	AA134737	Human colon cancer
24	14	56.0	273254	22	AA134737	Human colon cancer
25	14	56.0	349980	22	AA134737	Human colon cancer
26	13	52.0	281	22	AA134737	Human colon cancer
27	13	52.0	303	22	AA134737	Human colon cancer
28	13	52.0	351	22	AA134737	Human colon cancer
29	13	52.0	353	21	AA134737	Human colon cancer
30	13	52.0	375	21	AA134737	Human colon cancer
31	13	52.0	470	21	AA134737	Human colon cancer
32	13	52.0	480	22	AA134737	Human colon cancer
33	13	52.0	591	21	AA134737	Human colon cancer
34	13	52.0	606	21	AA134737	Human colon cancer
35	13	52.0	623	21	AA134737	Human colon cancer
36	13	52.0	996	21	AA134737	Human colon cancer
37	13	52.0	1081	22	AA134737	Human colon cancer
38	13	52.0	1114	21	AA134737	Human colon cancer
39	13	52.0	1284	19	AA134737	Human colon cancer
40	13	52.0	1284	19	AA134737	Human colon cancer
41	13	52.0	1359	21	AA134737	Human colon cancer
42	13	52.0	1443	21	AA134737	Human colon cancer
43	13	52.0	1497	21	AA134737	Human colon cancer
44	13	52.0	1566	22	AA134737	Human colon cancer
45	13	52.0	1782	21	AA134737	Human colon cancer
46	13	52.0	1783	21	AA134737	Human colon cancer
47	13	52.0	1812	21	AA134737	Human colon cancer
48	13	52.0	2022	22	AA134737	Human colon cancer
49	13	52.0	2046	22	AA134737	Human colon cancer
50	13	52.0	2114	22	AA134737	Human colon cancer
51	13	52.0	3119	22	AA134737	Human colon cancer
52	13	52.0	3127	22	AA134737	Human colon cancer
53	13	52.0	3164	22	AA134737	Human colon cancer
54	13	52.0	3172	21	AA134737	Human colon cancer
55	13	52.0	3926	21	AA134737	Human colon cancer
56	13	52.0	4145	21	AA134737	Human colon cancer
57	13	52.0	5255	21	AA134737	Human colon cancer
58	13	52.0	5520	19	AA134737	Human colon cancer
59	13	52.0	130480	22	AA134737	Human colon cancer
60	13	52.0	349980	22	AA134737	Human colon cancer
61	13	52.0	349980	22	AA134737	Human colon cancer
62	13	52.0	349980	22	AA134737	Human colon cancer
63	13	52.0	1038602	20	AA134737	Human colon cancer
64	12	48.0	19	14	AA051957	BCL-2 mRNA ribozyme
65	12	48.0	20	21	AA051957	BCL-2 mRNA ribozyme
66	12	48.0	50	14	AA050268	Dog genomic marker
67	12	48.0	61	16	AA050268	Dog genomic marker
68	12	48.0	61	16	AA050268	Dog genomic marker
69	12	48.0	110	19	AA129222	Family 2 bFGF DNA
70	12	48.0	110	19	AA129222	Family 2 bFGF DNA
71	12	48.0	110	19	AA129222	Family 2 bFGF DNA
72	12	48.0	110	19	AA129222	Family 2 bFGF DNA
73	12	48.0	110	19	AA129222	Family 2 bFGF DNA
74	12	48.0	110	19	AA129222	Family 2 bFGF DNA
75	12	48.0	110	19	AA129222	Family 2 bFGF DNA
76	12	48.0	110	19	AA129222	Family 2 bFGF DNA
77	12	48.0	110	19	AA129222	Family 2 bFGF DNA
78	12	48.0	110	19	AA129222	Family 2 bFGF DNA
79	12	48.0	110	19	AA129222	Family 2 bFGF DNA
80	12	48.0	110	19	AA129222	Family 2 bFGF DNA
81	12	48.0	110	19	AA129222	Family 2 bFGF DNA
82	12	48.0	110	19	AA129222	Family 2 bFGF DNA
83	12	48.0	110	19	AA129222	Family 2 bFGF DNA
84	12	48.0	110	19	AA129222	Family 2 bFGF DNA

Arabidopsis thalia
Glycine max isofla
C glutamicum codin
Corynebacterium g1
Mouse uncoupling p
Human DNA associat
Serine hydroxymeth
C. glutamicum lacti
Mechanical stress
Human DNA associat
Human cDNA encodin
Human genomic DNA
Chlamydia pneumoni
C. glutamicum codin
Human CLASP-7 intr
C glutamicum codin
Human secreted pro
Eucalyptus grandis
Human secreted pro
Human OREF ORF65
C glutamicum codin
Arabidopsis thalia
Arabidopsis thalia
Human OREF ORF752
Arabidopsis thalia
Cotton transcripti
Arabidopsis thalia
DNA encoding a S.
Streptococcus pneu
Arabidopsis thalia
C glutamicum codin
Arabidopsis thalia
C. glutamicum thalia
C. glutamicum SRT
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Nesleria gonorrhe
Human cDNA sequenc
Human cDNA sequenc
Human cDNA sequenc
Human polynucleoti
Human polynucleoti
Mulle 7-transmemb
Nucleotide sequenc
Nucleotide sequenc
DNA encoding de no
Human cDNA sequenc
Thermococcus 9N-2
R. marinus bacteri
C. glutamicum codin
C glutamicum codin
C glutamicum codin
Complete genome se
BCL-2 mRNA ribozyme
Dog genomic marker
HIV env INS mutage
Family 2 bFGF DNA
bFGF DNA ligand #5
Human biallelic po
Human biallelic po
Probe #125658 used
Probe #115163 for g
Probe #15188 used t
Probe #6930 used t
Eucalyptus grandis
Human TR13 coding
Human TR13 coding
Human TR13 coding
Mechanical stress
Human secreted pro
Probe #16821 for g
Probe #24370 used
Human gene express
Eucalyptus grandis
Staphylococcus aur

c	85	12	48.0	387	21	AAAB57407	Eucalyptus grandis
c	86	12	48.0	393	20	AAAB06347	Kidney injury assoc
c	87	12	48.0	393	21	AAAC74732	Human OREX ORE287
c	88	12	48.0	400	18	AAAV78325	Staphylococcus aure
c	89	12	48.0	400	18	AAAV78307	Staphylococcus aur
c	90	12	48.0	402	14	AAO59673	Human brain Express
c	91	12	48.0	408	21	AAAB74402	Eucalyptus grandis
c	92	12	48.0	419	20	AAAB75566	Extended human seq
c	93	12	48.0	424	21	AAAB74409	Eucalyptus grandis
c	94	12	48.0	427	21	AAFO94042	Fusarium venenatum
c	95	12	48.0	429	22	AAAB87991	Peppermint plant c
c	96	12	48.0	433	21	AAAC98600	Human colon cancer
c	97	12	48.0	433	21	AAAB74415	Eucalyptus grandis
c	98	12	48.0	434	21	AAAC45655	Arabidopsis thalia
c	99	12	48.0	445	21	AAAC08146	Human secreted pro
c	100	12	48.0	451	22	AAI627860	Human CDNA seq ID

ALIGNMENTS

RESULT	1
XX	AAI62941
ID	AAI62941 standard; DNA: 839 BP.
XX	
AC	AAI62941;
XX	
DT	22-OCT-2001 (first entry)
XX	
DE	Human genomic DNA SEQ ID NO 269.
XX	
XX	Human: neutropenic; neuroprotective; cytosstatic; dermatological; virucide
XX	immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerrary
XX	antiparkinsonian; antislackling; antianaemic; antiarthritic; cancer;
XX	antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
XX	antiallergic; antidiabetic; anticulcer; anticonvulsant; antifungal;
XX	antiparasitic; cardiant; immune disorder; cardiovascular disorder;
XX	neurological disease; infection; nephrotropic; gene therapy; vaccine;
XX	ds.
XX	
OS	Homo sapiens.
XX	
PN	WO20015449-A1.
XX	
PD	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01346.
XX	
PR	31-JAN-2000; 2000US-0179065.
XX	
PR	04-FEB-2000; 2000US-0180628.
XX	
PR	19-MAY-2000; 2000US-0205515.
XX	
PR	07-JUL-2000; 2000US-0216880.
XX	
PR	14-JUL-2000; 2000US-0218290.
XX	
PR	14-AUG-2000; 2000US-0225447.
XX	
PR	01-SEP-2000; 2000US-0229343.
XX	
PR	06-SEP-2000; 2000US-0230437.
XX	
PR	08-SEP-2000; 2000US-0231243.
XX	
PR	25-SEP-2000; 2000US-0234997.
XX	
PR	29-SEP-2000; 2000US-0236367.
XX	
PR	13-OCT-2000; 2000US-0239937.
XX	
PR	08-NOV-2000; 2000US-0246476.
XX	
PR	08-NOV-2000; 2000US-0246477.
XX	
PR	08-NOV-2000; 2000US-0246525.
XX	
PR	08-NOV-2000; 2000US-0246526.
XX	
PR	08-NOV-2000; 2000US-0246528.
XX	
PR	17-NOV-2000; 2000US-0246528.
XX	
PR	17-NOV-2000; 2000US-0249210.
XX	
PR	17-NOV-2000; 2000US-0249211.
XX	
PR	17-NOV-2000; 2000US-0249214.
XX	
PR	17-NOV-2000; 2000US-0249265.
XX	
PR	01-DEC-2000; 2000US-0250160.
XX	
PR	01-DEC-2000; 2000US-0250391.
XX	
PR	05-DEC-2000; 2000US-0251030.
XX	
PR	05-DEC-2000; 2000US-0251988.

05-DEC-2000; 2000US-0256719.
06-DEC-2000; 2000US-0251479.
08-DEC-2000; 2000US-0251989.
08-DEC-2000; 2000US-0251990.
11-DEC-2000; 2000US-0254097.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-476225/51.
Novel plasma membrane associated proteins useful for diagnosing,
treating, preventing and/or prognosing disorders related to the
proteins, including cancer, immune response and neuronal disorders
Example 2: SEQ ID NO 269; 532bp + Sequence Listing; English.

The invention relates to novel genes (AA162572, AA162961) and proteins (AA442447-AA442415) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. bone and ovarian cancer and other cancers of the adrenal gland, bone, prostate, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WHO at ftp.who.int/pub/published_pat_sequences.

Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 Other;

Query Match	100.0%;	Score 25;	DB 22;	length 839;
Best Local Similarity	100.0%;	Pred. No. 1.2e-05;		
Matches 25;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

```
QY      1   gcatctcgatcctcgtcaggtgcag    25  
          |||||  
Db     131   gcatttcgatcctcgtcagtgcag    155
```

	RESULT	2
XX	AAx01303	
ID	AAx01303 standard; DNA; 1041 BP.	
XX		
AC	AAx01303;	
XX		
DT	12-APR-1999 (first entry)	
XX		
DE	E. coli biotin synthetase (Biot) coding sequence.	
XX		
KW	DAP aminotransferase; diaminopelargonic acid; transgenic plant,	
KM	biotin synthase; biotin production; vitamin H; Biot; ss.	
XX		
OS	Escherichia coli.	
PN	US5869719-A.	
PD		
XX	09-FEB-1999.	
PF		
XX	30-APR-1997; 97US-0846338.	
PR	30-APR-1997; 97US-0846338.	
XX	08-MAR-1995; 95US-0401068.	
XX		

PA (NOVS) NOVARTIS FINANCE CORP.
 XX Patton DA.
 XX WPI: 1999-152902/13.
 DR P-PSDB; AAW73906.
 XX
 PT Transgenic plants with high biotin levels - transformed with DNA
 encoding di:amino-pelargonic acid amino-transferase or biotin
 synthase
 PT
 XX
 PS Example 2; Column 37-40; 34pp; English.
 CC This sequence encodes the E. coli biotin synthetase (BIOB). The gene can
 be used in the transgenic plant of the invention. The transgenic plant,
 plant cell or plant tissue is transformed with a chimeric gene encoding
 diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 humans or animals.
 CC
 XX
 SQ Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcattcgatccctgcagtgacg 25
 ||||||||||||||||||||
 Db 99 gcattcgatccctgcagtgacg 123

RESULT 3
 AAN91329
 ID AAN91329 standard; DNA; 1084 BP.
 XX
 AC AAN91329;
 XX
 DT 15-FEB-1990 (first entry)
 XX
 DE E.coli Bio B gene.
 XX
 KW E.coli Bio B gene; biotin.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT CDS 24..1064
 FT /*tag=a
 XX
 PM GB2216530-A.
 XX
 PD 11-OCT-1989.
 XX
 PF 17-MAR-1989; 89GB-0006210.
 XX
 PR 22-MAR-1988; 88GB-0006804.
 PR 17-MAR-1989; 89GB-0006210.
 XX
 PA (UKAG-) UK MTN. AGRIC. FISH.
 XX
 PI Pearson BM, McKee RA;
 XX
 DR WPI: 1989-295085/41. P-PSDB 1-1392
 XX
 PT Plasmid contg. gene(s) for ex: assion of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.
 XX
 PS Table 3; page 33-4; 52pp; English.
 XX
 CC The gene can be used in a plasmid for expression of enzymes of the biotin

CC synthetic pathway. Pref. control sequences for expression in *S.cerevisiae*
 CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
 CC *Lactobacillus*. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.
 XX
 SQ Sequence 1084 BP; 271 A; 266 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcattcgatccctgcagtgacg 25
 ||||||||||||||||||||
 Db 122 gcattcgatccctgcagtgacg 146

RESULT 4
 AAN60496
 ID AAN60496 standard; DNA; 1121 BP.
 XX
 AC AAN60496;
 XX
 DT 17-OCT-1991 (first entry)
 XX
 DE Sequence encoding biotin synthesising enzyme.
 XX
 KW Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
 XX
 FH Key Location/Qualifiers
 FT CDS 42..1082
 FT /*tag=a
 XX
 PN JP61149091-A.
 XX
 PD 07-JUL-1986.
 XX
 PF 24-DEC-1984; 84JP-0272605.
 XX
 PR 24-DEC-1984; 84JP-0272605.
 XX
 PA (NIPS) NIPPON SODA KK.
 XX
 DR WPI: 1986-216622/33.
 DR P-PSDB; AAP60536.
 XX
 PT Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
 XX
 PS Disclosure; Page 534; 23pp; Japanese.
 XX
 CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing desthiobiotin.
 XX
 SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcattcgatccctgcagtgacg 25
 ||||||||||||||||||||
 Db 140 gcattcgatccctgcagtgacg 164

RESULT 5
 AA062386
 ID AA062386 standard; DNA; 5872 BP.
 XX

```

AC AA062386;
XX
XX 16-NOV-1994 (first entry)
DE Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
XX Biotin: expression: enterobacteria; vitamin H; synthesis:
KM plasmid: pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
KM Promoter plac; biotin synthase; KAPA synthase;
KM 8-aminio-7-oxononanoate synthase; pimeloyl-CoA; DFB synthase;
KM dehydrobiotin synthase; DAPA synthase;
KM S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
KM seborrhoea; dermatitis; ds.
XX
XX Escherichia coli DSM498.
XX
XX Key
FH Location/Qualifiers
FH promoter
FT /tag= a
FT /function= "promoter plac"
FT /evidence= EXPERIMENTAL
FT
FT -35_signal
FT /tag= b
FT /standard_name= "promoter plac"
FT /evidence= c
FT /evidence= EXPERIMENTAL
FT /standard_name= "promoter plac"
FT /tag= d
FT /evidence= EXPERIMENTAL
FT /standard_name= "promoter plac"
FT
FT -10_signal
FT /tag= e
FT /product= "biotin synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioB"
FT /number= 1
FT 1141..1146
FT /tag= f
FT /standard_name= "bioF RBS"
FT 1154..2311
FT /tag= g
FT /EC_number= 2.3.1.47
FT /product= "KAPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioF"
FT
FT CDS
FT /number= 2
FT /standard_name= "8-amino-7-oxononanoate synthase"
FT 2284..2288
FT /tag= h
FT /standard_name= "bioc RBS"
FT 2295..3050
FT /tag= i
FT /function= "involved in pimeloyl-CoA synthesis"
FT /product= "protein"
FT /gene= "bioc"
FT /number= 3
FT 3030..3033
FT /tag= j
FT /standard_name= "bioc RBS"
FT 3043..3753
FT /tag= k
FT /EC_number= 6.3.3.3
FT /product= "DPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "biocD15"
FT /number= 4
FT /standard_name= "dehydrobiotin synthase"
FT 3712..3750
FT /tag= l
FT /note= "biocD15 substitution"
FT 3742..3746
FT RBS

```

```

FT /tag= m
FT /standard_name= "bioA RBS"
FT 3750..5039
FT /tag= n
FT /EC_number= 2.6.1.62
FT /product= "DAPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioA"
FT /number= 5
FT /standard_name= "S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase"
FT 5088..5093
FT /tag= o
FT /standard_name= "ORF1 RBS"
FT 5098..5574
FT CDS
FT /tag= p
FT /function= "unknown, involved in biotin synthesis"
FT /product= "protein"
FT /evidence= EXPERIMENTAL
FT /gene= "ORF1"
FT /number= 6
FT 5583..5644
FT /tag= q
FT /standard_name= "rho-independent transcriptional terminator"
FT 5583..5605
FT /tag= r
FT
FT stem_loop
FT 5583..5605
FT /tag= r
FT
FT WO9408023-A.
FT
FT 14-APR-1994.
FT
FT 01-OCT-1993; 93WO-EP02688.
FT
FT 02-OCT-1992; 92CH-0003124.
FT 15-JUL-1993; 93CH-0002134.
FT
FT (LONZ ) LONZA AG.
FT
FT Birch O, Brass J, Fuhrmann M, Shaw N;
FT WPT: 1994-135587/16.
FT DR P-PADB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
FT
FT Biotechnological biotin prodn. using enterobacterial biotin-gene
FT - providing vitamin H in high yield
FT
FT Claim 1, Fig 6, page 47-55 and 60-65; 92pp; German.
FT
FT The sequence is derived from plasmid pB030A-15/9 contg. the
FT bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
FT of biotin, arranged in a transcription unit. Microorganisms
FT contg. these DNA fragments or plasmids may be used in the prodn.
FT of biotin. Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
FT loss of appetite and tiredness.
FT
FT SO Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;

```

Query Match 100.0%; Score 25; DB 15; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 1.1e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gatttcgactctgcagtgacg 25
 |||
 Db 215 gatttcgactctgcagtgacg 239

RESULT 6
 AA075832
 ID AA075832 standard: DNA: 34 BP.
 XX
 AC AA075832;

XX 18-AUG-1995 (first entry)
 DT Sense primer to amplify Non-A Non-B hepatitis virus for analysis.
 XX
 XX Non-A Non-B hepatitis virus: structural region: cDNA to genomic RNA:
 DE detection; reagent; anti-Non-A Non-B hepatitis virus antibody;
 KW vaccine; antigen; epitope; diagnosis; ss.
 XX
 XX Synthetic.
 OS
 XX
 XX EP628572-A.
 PN
 XX 14-DEC-1994.
 PD
 XX 27-MAY-1994; 94EP-0108256.
 PF
 XX 28-MAY-1993; 93JP-0126709.
 PR 02-MAR-1994; 94JP-0032201.
 PR
 XX (ARIM/) ARIMA T.
 PA (EISA) EISAI CO LTD.
 PA
 XX Aoyama M, Arima T, Hosoda T, Iwasaki Y, Obara T;
 PI Sawada T, Tohmatsu J;
 PI WPI, 1995-015655/03.
 DR
 XX New non-A non-B hepatitis virus sub-type - used to develop prods.
 PT for detection, diagnosis, prevention and treatment of non-A non-B
 PT hepatitis.
 PT
 XX Example 2; Page 49; 59pp; English.
 PS
 XX This primer is based on nucleotides 778-800 of AA075818 (a part of
 CC the Non-A Non-B hepatitis virus genome encoding the non-structural
 CC protein). It is used in conjunction with AA075833 to amplify nucleotides
 CC 800-1336 of AA075818. The nucleotide sequences (see also AA075817-19)
 CC were isolated from the plasma of donors in Japan with high s-GTP levels,
 CC and were found to be different from previously reported NANB hepatitis
 CC viruses. The DNA can be used as a reagent for detecting the NANB
 CC hepatitis viral gene. The polypeptides can be used as reagents for
 CC detecting anti-NANB hepatitis antibodies or as a NANB hepatitis viral
 CC vaccine.
 CC
 XX Sequence 34 BP; 6 A; 9 C; 12 G; 7 T; 0 other;
 SQ
 Query Match 60.0%; Score 15; DB 16; Length 34;
 Best Local Similarity 100.0%; Pred. No. 8.9;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 8 gatcctgcagtg 22
 Db 8 gatcctgcagtg 22
 OY
 Db
 RESULT 7
 AAC75563/C
 ID AAC75563 standard; cDNA: 586 BP.
 XX
 XX AAC75563;
 AC
 XX 08-FEB-2001 (first entry)
 DT
 XX Human ORFX ORF118 polynucleotide sequence SEQ ID NO:2235.
 DE
 XX Human: open reading frame: ORFX: detection; cytosstatic; hepatotropic;
 KW vulnery; antiprositic; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;
 KW antiviral; antibacterial; antifungal; antineumatic; antihypoid;

KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX
 XX Homo sapiens.
 OS
 XX
 XX WO200058473-A2.
 PN
 XX 05-OCT-2000.
 PD
 XX 31-MAR-2000; 2000WO-US08621.
 PF
 XX 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 PR
 XX (CURA-) CURAGEN CORP.
 PA
 XX Slimkets RA, Leach M;
 PI WPI: 2000-602362/57.
 DR P-PSDB; AAB41354.
 DR
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 PT
 XX Claim 5; Page 1649; 5507pp; English.
 PS
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytosstatic; hepatotropic; vulnery;
 CC antiprositic; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antinflammatory; antibacterial; antiviral; antifungal; antineumatic;
 CC antihypoid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 XX Sequence 586 BP; 122 A; 185 C; 156 G; 123 T; 0 other;
 SQ
 Query Match 60.0%; Score 15; DB 21; Length 586;
 Best Local Similarity 100.0%; Pred. No. 7.6;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 9 atcctgcagtg 23
 Db 304 ATCCCTGCTCAGGTGC 290
 OY
 Db
 RESULT 8
 AAH43851/C
 ID AAH43851 standard; cDNA: 6372 BP.
 XX

AC	AAH43851;	
XX		
DT	04-SEP-2001	(first entry)
XX		
DE	Human CLASP-7 encoding cDNA sequence SEQ ID NO:1.	
XX		
KW	Human; CLASP-7; cadherin-like asymmetry protein; immune response;	
KW	neuroprotective; antidiabetic; immunosuppressive; antirheumatic;	
KW	antiarthritic; hypotensive; anti-HIV; cytostatic; immunostimulant;	
KW	antianemic; antinflammatory; ophthalmological; nephrotrophic;	
KW	antihypertoid; antiaslathmic; antiallergic; antibacterial; gene therapy;	
KW	chromosome 19q13.2; autoimmune disease; multiple sclerosis; toxemia;	
KW	juvenile diabetes; rheumatoid arthritis; pruritic urticarial papule;	
KW	hypertension; Rh incompatibility; ss.	
XX		
OS	Homo sapiens.	
FT	Key	Location/Qualifiers
FT	CDS	13..6156
FT		/*tag= a
FT		/product= "CLASP-7"
XX		
XX	WO200142295-A2.	
PD	14-JUN-2001.	
XX		
PF	13-DEC-2000; 2000WO-US34152.	
XX		
PR	13-DEC-1999; 99US-0170453.	
PR	14-JAN-2000; 2000US-0176195.	
PR	14-FEB-2000; 2000US-0182286.	
PR	11-APR-2000; 2000US-0196287.	
PR	11-APR-2000; 2000US-0196527.	
PR	11-APR-2000; 2000US-0196528.	
PR	11-APR-2000; 2000US-0196528.	
PR	13-OCT-2000; 2000US-0240503.	
PR	13-OCT-2000; 2000US-0240508.	
PR	13-OCT-2000; 2000US-0240539.	
PR	13-OCT-2000; 2000US-0240543.	
XX		
PA	(ARBO-) ARBOR VITA CORP.	
PI	Lu P, Garman JD, Candia AF;	
XX		
XX	WPI: 2001-361641/40.	
DR	P-PSDB: AAB99541.	
XX		
PT	Novel cadherin-like asymmetry protein-7 and polynucleotides encoding	
PT	the polypeptide, useful for treating autoimmune disease,	
PT	hypersensitivity, preventing transplant rejection by modulating immune	
PT	response -	
XX		
PS	Claim 1; Fig 5; 151pp; English.	
XX		
CC	The present invention describes a human cadherin-like asymmetry protein	
CC	(CLASP), designated CLASP-7. The CLASP-7 protein (I) and its encoding	
CC	nucleotide sequence (II) have activities including: neuroprotective;	
CC	antidiabetic; immunosuppressive; antirheumatic; anti-HIV;	
CC	hypotensive; cytostatic; immunostimulant; antianemic; antinflammatory;	
CC	ophthalmological; nephrotrophic; antihypertoid; antiaslathmic;	
CC	antiallergic; and antibacterial. (II) and CLASP-7 antibodies (III) are	
CC	useful for detecting the CLASP-7 polypeptide. (II) is useful for	
CC	producing (I) by recombinant methods. (I) or its fragment are useful for	
CC	inhibiting an immune response in a cell such as T cell or B cell. A	
CC	pharmaceutical composition (C), comprising (I) or (II), can be useful	
CC	for treating CLASP-7-mediated disease such as an autoimmune disease	
CC	caused or exacerbated by increased activity of T helper cells. Autoimmune	
CC	diseases which can be treated using (C) include multiple sclerosis,	
CC	juvenile diabetes and rheumatoid arthritis. (I) is useful for treating	
CC	toxemia or pregnancy induced hypertension, pruritic urticarial papules	
CC	and Rh incompatibility. (I) is also useful as a diagnostic reagent for	
CC	immune and other disorders, since diseases characterised by	

```

CC overproduction or depletion of lymphocytes in blood or other organs may
CC be detected by monitoring the level of (1) or its mRNA. CLASP-7 has been
CC mapped to the chromosomal location 19q13.2. The present sequence
CC encodes the human CLASP-7 protein as given in the present invention.
XX
SQ Sequence 6372 BP: 1273 A; 2050 C; 1849 G; 1200 T; 0 other:
Query Match 60.0%; Score 15; DB 22; Length 6372;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY 9 atcctgcagcagtc 23
|||||
DB 735 ATCCCGCTCAGGTGC 721
RESULT 9
AAZ19643
ID AAZ19643 standard; RNA: 66 BP.
XX
AC AAZ19643;
XX
DT 08-NOV-1999 (first entry)
XX
DE Complement system protein C1q RNA binding ligand 29.
XX
KW Complement system protein; C1q; C3; RNA ligand; neurotropic; stroke;
KW neuroprotective; anti-asthmatic; cerebroprotective; immunosuppressive;
KW diagnostic; prevention; treatment; complement protein-related disease;
KW Alzheimer's disease; renal disease; transplant rejection; asthma; ss.
XX
OS Synthetic.
XX
PN WC0941271-A1.
XX
PD 19-AUG-1999.
XX
PE 05-FEB-1999; 99MO-US02597.
XX
PR 29-SEP-1998; 98US-0163025.
XX
PR 12-FEB-1998; 98US-0023228.
XX
PA (NEXS-) NEXSTAR PHARM INC.
XX
PI Biesecker G, Gold L;
XX
DR WPI: 1999-527357/44.
XX
PT New Nucleic Acid ligand to complement protein C5, useful in the
PT treatment of Alzheimer's disease, renal diseases, transplant
PT rejection, stroke and asthma
XX
PS Example 6; Page 45; 120bp; English.
XX
CC This invention describes novel purified and isolated non-naturally
CC occurring RNA ligands (1) to complement system proteins C5, C1q and C3.
CC The products of the invention have neurotropic, neuroprotective,
CC anti-asthmatic, cerebroprotective and immunosuppressive activity. The
CC new ligands are useful as diagnostic agents, and pharmaceutical agents
CC for prevention and treatment of complement protein-related diseases,
CC including Alzheimer's disease, renal diseases, transplant rejection,
CC stroke and asthma.
XX
SQ Sequence 66 BP: 11 A; 19 C; 23 G; 13 U; 0 other:
Query Match 56.0%; Score 14; DB 20; Length 66;
Best Local Similarity 71.4%; Pred. No. 33;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0
YY 6 tcgattcctcgtaag 19
|||||

```


Db 40 ucgaucucgucag 53

RESULT 10

AAH34737/C standard; cDNA: 370 BP.

AAH34737:

03-SEP-2001 (first entry)

Human colon cancer antigen encoding cDNA SEQ ID NO:1819.

Human: colon cancer; colon cancer antigen; diagnosis; detection; colorectal carcinoma; ss.

Homo sapiens.

MO200122920-A2.

05-APR-2001.

28-SEP-2000; 2000MO-US26524.

29-SEP-1999; 99US-0157137.

03-NOV-1999; 99US-0163280.

(HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Barash SC, Birse CE, Rosen CA;

WPI: 2001-235357/24.

P-PSDB; AAG75332.

Nucleic acids encoding 4277 human colon cancer-associated polypeptides, useful for preventing, diagnosing and/or treating colorectal cancers -

Claim 1; Page 3364; 9803pp: English.

AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patient's own production of P. Additionally, N may be used to produce the colon cancer-associated P, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent sequences used in the exemplification of the present invention.

N.B. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1027 to 1052, 7921 and 7922.

Sequence 370 BP; 112 A; 55 C; 81 G; 119 T; 3 other;

Query Match 56.0%; Score 14; DB 22; Length 370;

Best Local Similarity 100.0%; Pred. No. 29; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 12 ctgcgtacgtgcag 25

DB 317 ctgcgtacgtgcag 304

RESULT 11

AAC35170

ID AAC35170 standard; DNA: 974 BP.

AAC35170:

17-OCT-2000 (first entry)

Arabidopsis thaliana DNA fragment SEQ ID NO: 9244.

Hybridisation assay; genetic mapping; gene expression control; protein identification; signal transduction pathway; metabolic pathway; promoter; termination sequence; ss.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

05-MAR-1999; 99US-0123180.

09-MAR-1999; 99US-0123548.

23-MAR-1999; 99US-0125788.

25-MAR-1999; 99US-0126264.

29-MAR-1999; 99US-0126785.

01-APR-1999; 99US-0127462.

06-APR-1999; 99US-0128234.

08-APR-1999; 99US-0128714.

16-APR-1999; 99US-0129845.

19-APR-1999; 99US-0130077.

21-APR-1999; 99US-0130449.

23-APR-1999; 99US-0130510.

28-APR-1999; 99US-0130891.

30-APR-1999; 99US-0131449.

30-APR-1999; 99US-0132048.

04-MAY-1999; 99US-0132407.

05-MAY-1999; 99US-0132484.

06-MAY-1999; 99US-0132485.

06-MAY-1999; 99US-0132486.

06-MAY-1999; 99US-0132487.

07-MAY-1999; 99US-0132863.

11-MAY-1999; 99US-0134256.

14-MAY-1999; 99US-0134218.

14-MAY-1999; 99US-0134219.

14-MAY-1999; 99US-0134221.

14-MAY-1999; 99US-0134768.

18-MAY-1999; 99US-0134941.

20-MAY-1999; 99US-0135124.

21-MAY-1999; 99US-0135353.

24-MAY-1999; 99US-0135629.

25-MAY-1999; 99US-0136021.

27-MAY-1999; 99US-0136392.

28-MAY-1999; 99US-0136782.

01-JUN-1999; 99US-0137222.

03-JUN-1999; 99US-0137528.

04-JUN-1999; 99US-0137502.

07-JUN-1999; 99US-0137724.

08-JUN-1999; 99US-0138094.

10-JUN-1999; 99US-0138540.

10-JUN-1999; 99US-0138847.

14-JUN-1999; 99US-0139119.

16-JUN-1999; 99US-0139452.

16-JUN-1999; 99US-0139453.

17-JUN-1999; 99US-0139492.

18-JUN-1999; 99US-0139454.

18-JUN-1999; 99US-0139455.

18-JUN-1999; 99US-0139456.

18-JUN-1999; 99US-0139457.

18-JUN-1999; 99US-0139458.

18-JUN-1999; 99US-0139459.

18-JUN-1999; 99US-0139460.

Query Match	Score 14:	DB 21:	Length 974:
Best Local Similarity	100.0%	Pred. No. 28:	
Matches 14:	Conservative	0:	Mismatches 0: Indels 0: Gaps 0:
Oy	4	tttcgacccctgc 17	
Db	593	tttcgacccctgc 606	
RESULT 12			
AAC48947			
XX	AAC48947 standard; DNM: 974 BP.		
AC	AAC48947:		
XX	18-OCT-2000 (first entry)		
DT			
XX			

XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
OS EPI033405-A2.
PN
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134376.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.

```

PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157733.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0158295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 56.0%; Score 14; DB 21; Length 974;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 4 ttgcgtctctgc 17
    |||
Db 593 ttgcgtctctgc 606

```

RESULT 13
AAA29322

ID AAA29322 standard; cDNA; 1160 BP.

XX AAA29322;

XX 26-SEP-2000 (first entry)

DE Glycine max isoflavone O-methyltransferase clone srl.pK0015.t4.

KW Isoflavone O-methyltransferase; soybean; degradation; daidzein;

KW Flavonoid; flower colour; growth; pollination; irradiation; ss.

OS Glycine max.

XX Key Location/Qualifiers

```

FF CDS 11..1087
FT /*lag= a
FT /product= Isoflavone_O-methyltransferase
XX

```

PN W0200037656-A2.

XX 29-JUN-2000.

PF 20-DEC-1999; 99WO-US30338.

XX 21-DEC-1998; 98US-0113190.

PA (DUPO) DU PONT DE NEMOURS & CO E I.

PI Famodu OO, McConigle B, Odell JT, Fader GM, Falco SC;

DR WPI: 2000-442680/38.

XX P-PSDB; AA96584.

PT New polynucleotide encoding flavonoid biosynthetic enzymes for creating
PS transgenic plants and for immunological screening of cDNA libraries
XX Claim 2; Page 32; 39pp; English.

CC AAA29321-25 encode isoflavone O-methyltransferases isolated from various
CC soybean (Glycine max) tissues. cDNA libraries set4, srl, stric and sllc
CC were prepared from soybean embryo (19 days after flowering), root, 8-day
CC old root and seed (25 days after flowering).
CC Isoflavone O-methyltransferase catalyses the first step in degradation
CC of daidzein. Suppression of this enzyme will yield higher concentrations
CC of this beneficial isoflavone in, e.g., soybean seed. Flavonoids serve as
CC co-pigments in flower colour, stimulate pollen tube growth, attract
CC pollinators, act as feeding deterrents and protect against UV
CC irradiation in fruits and seeds. The cDNA and proteins can be used to
CC isolate homologues, for immunological screening and for positive
CC selection methods.

SO Sequence 1160 BP; 361 A; 197 C; 254 G; 348 T; 0 other;

```

Query Match 56.0%; Score 14; DB 21; Length 1160;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 12 ctgcgcaggtgcag 25
    |||
Db 222 ctgcgcaggtgcag 235

```

RESULT 14

ID AAH66075 standard; DNA; 1302 BP.

XX AAH66075;

XX 26-SEP-2001 (first entry)

DE C glutamicum coding sequence fragment SEQ ID NO: 1110.

KW Corynebacterium bacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis; ds.

OS Corynebacterium glutamicum.

XX EPI108790-A2.

PD 20-JUN-2001.

PF 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOWA) KYOWA HAKKO KOSYO KK.
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI: 2001-376931/40.
DR P-PSDB: AAG90856.
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
PS Claim 8: SEQ ID NO: 1110; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Corynebacterium, and identifying a homologue of a gene derived
CC from Corynebacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 1302 BP; 241 A; 326 C; 374 G; 361 T; 0 other:

Query Match 56.0%; Score 14; DB 22; Length 1302;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ttctgacctctc 17
DB 1080 ttctgacctctc 1093

RESULT 15
AA71824
ID AAF71824 standard; DNA; 1425 BP.
XX
AC AAF71824;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:143.
XX
KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW fine chemical production; microorganism; organic acid; nucleoside;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100843-A2.
XX
PD 04-JAN-2001.
XX
PE 23-JUN-2000; 2000MO-1B00923.
XX
XX 25-JUN-1999; 9905-0141031.
PR 01-JUL-1999; 99DE-1030476.
PR 02-JUL-1999; 9905-0142101.
PR 08-JUL-1999; 99DE-1031415.
PR 08-JUL-1999; 99DE-1031416.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.

PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031435.
PR 08-JUL-1999; 99DE-1031443.
PR 08-JUL-1999; 99DE-1031453.
PR 08-JUL-1999; 99DE-1031457.
PR 08-JUL-1999; 99DE-1031465.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031541.
PR 08-JUL-1999; 99DE-1031573.
PR 08-JUL-1999; 99DE-1031592.
PR 08-JUL-1999; 99DE-1031632.
PR 08-JUL-1999; 99DE-1031634.
PR 08-JUL-1999; 99DE-1031636.
PR 08-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032126.
PR 09-JUL-1999; 99DE-1032130.
PR 09-JUL-1999; 99DE-1032186.
PR 09-JUL-1999; 99DE-1032206.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032922.
PR 14-JUL-1999; 99DE-1032926.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1033004.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 12-AUG-1999; 9905-0148613.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040832.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041380.
PR 31-AUG-1999; 99DE-1041394.
PR 31-AUG-1999; 99DE-1041396.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 09-MAR-2000; 2000US-0187970.
XX
PA (BADI) BASF AG.
XX
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
PI WPI: 2001-137957/14.
DR P-PSDB: AAB79705.
XX
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
PT pathway proteins, useful for producing fine chemicals in
PT microorganisms, including organic acids, nonproteinogenic amino acids,
PT and purine and pyrimidine bases -
XX
PS Claim 3; Page 374-377; 1737pp; English.
XX
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
CC MP nucleic acids are useful for the production of fine chemicals
CC in microorganisms, including organic acids, nonproteinogenic amino
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
CC compounds, vitamins, cofactors, polyketides and enzymes.

XX Sequence 1425 BP; 260 A; 348 C; 410 G; 407 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1425;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 ttctgactctcgtc 17
Db 1180 ttctgactctcgtc 1193

RESULT 16
AAV44597/c
ID AAV44597 standard; DNA; 1758 BP.

XX AAV44597;
XX
XX
XX 24-NOV-1998 (first entry)

DE Mouse uncoupling protein-2 UCP2 gene promoter region.

XX Uncoupling protein-2; UCP2 gene; mouse; respiration;
XX thermogenesis; obesity; hyperinsulinemia; glucose intolerance;
XX diabetes; syndrome X; hypothermia; wasting; cachexia; anorexia;
XX inflammation; fever; hyperthermia; gene therapy; diagnosis; ds.

OS Mus sp.

XX W09831396-A1.

XX 23-JUL-1998.

XX 22-APR-1997; 97WO-US06864.

XX 15-JAN-1997; 97US-0034960.

PA (NARE-) CENT NAT RECH SCI CENT RECH SUR ENDOCRINOL.

PA (RECC) UNIV CALIFORNIA.

XX (UTDU-) UNIV DUKE.

PI Boullaud F, Collins SA, Ricquier D, Seidlin MF;

PI Surwit RS, Warden CH;

XX WPI: 1998-413823/35.

PT Method for treating disease associated with altered UCP-2 expression
PT - by administering agent which enhances or inhibits UCP-2 activity,
PT cachexia etc.

PS Example IT: Fig 8a; 98pp; English.

XX This is the nucleotide sequence of a fragment of the mouse
CC uncoupling protein-2 (UCP2) gene, corresponding to the promoter
CC region, obtained from genomic clone MMU2-12. A DNA fragment from
CC -934 to +8600 bp of the UCP2 gene has also been sequenced (see
CC AAV44596). MMU2-12 (deposited as T-1868) was cloned from a mouse
CC genomic library by screening with mouse ucp2 cDNA; the cDNA had
CC been obtained from a mouse muscle cDNA library using rat UCP1 cDNA
CC as probe. Human UCP2 cDNA and genomic sequences (see AAV44595,
CC AAV44598-602 and AAV44629-38) are also provided, as well as a deduced
CC amino acid sequence (see AAW69166) for human UCP2. The invention
CC relates to a method for treating disease associated with altered
CC UCP2 expression. An agent which enhances UCP2 expression can be
CC used to treat obesity, diabetes, syndrome X, hypothermia,
CC hyperinsulinemia, or glucose intolerance. An inhibitor of UCP2
CC is used to treat wasting, anorexia, inflammation, cachexia, fever
CC or hyperthermia. The invention also relates to diagnostic and drug
CC screening methodologies.

XX Sequence 1758 BP; 421 A; 401 C; 438 G; 487 T; 11 other;

Query Match 56.0%; Score 14; DB 19; Length 1758;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 atcctcgtcaggtg 22
Db 1072 ATCCTCGTCAGGTG 1059

RESULT 17
AAS07381/c
ID AAS07381 standard; cDNA; 2014 BP.

XX AAS07381;
XX
XX
XX 26-SEP-2001 (first entry)

DE Human DNA associated with CLASP-5 #1.

XX Human; CLASP-5; cadherin-like asymmetry protein; immune gateway;
XX immunogen; antibody; autoimmune disease; rheumatoid arthritis;
XX multiple sclerosis; leukaemia; insulin dependent diabetes mellitus;
XX acquired immunodeficiency syndrome; AIDS; ss.

OS Homo sapiens.

XX location/Qualifiers

FT Key

FT CDS

FT 2..2014
FT /*tag= a
FT /product= "Protein as displayed in AAU04025"
FT /partial
FT /note= "No start codon"
FT /transl_except= (pos:1523..1525 aa:Arg)
FT /transl_except= (pos:1883..1885 aa:Ile)

XX W0200142296-A2.

XX 14-JUN-2001.

XX 13-DEC-2000; 2000WO-US34163.

XX 13-DEC-1999; 99US-0170453.

XX 14-JAN-2000; 2000US-0176195.

XX 14-FEB-2000; 2000US-0182296.

XX 11-APR-2000; 2000US-0196267.

XX 11-APR-2000; 2000US-0196460.

XX 11-APR-2000; 2000US-0196527.

XX 11-APR-2000; 2000US-0196528.

XX 13-APR-2000; 2000US-0547276.

XX 13-OCT-2000; 2000US-0240503.

XX 13-OCT-2000; 2000US-0240508.

XX 13-OCT-2000; 2000US-0240543.

XX 13-OCT-2000; 2000US-0240539.

XX (ARBO-) ARBOR VITA CORP.

XX Lu P, Garman JD, Candia AF;

XX WPI: 2001-367865/38.

XX P-PsDB; AAU04025.

XX CLASP-5 polynucleotides, proteins and antibodies are used to prevent or

PT treat a CLASP-5 mediated disease, such as an autoimmune disease e.g.

PT Rheumatoid arthritis -

PS Disclosure; Fig 9A; 188pp; English.

XX The sequence encodes a protein presented as AAU04025, included in the

CC specification which relates to a cadherin-like asymmetry protein,

CC CLASP-5, which is a transmembrane protein of the immune system involved

in the formation of the immune gateway. CLASP-5, polynucleotides encoding

CC sense mutant of (S2) or (S3). (A), particularly L-threonine and
 CC L-isoleucine, are useful in human medicine, in the food and
 CC pharmaceutical industries, and especially in animal nutrition. Bacteria
 CC with reduced glyla activity produce higher yields of (A). The strain
 CC *Brevibacterium flavum* DM368-2/pZI produced, after 72 hr culture, 1.32 g/l
 CC L-threonine. The strain DM368-2::pK18mobglyA', in which the glyla gene had
 CC been suppressed by integrational mutagenesis, produced 1.60 g/l. This
 CC sequence represents the lact-tac-glyla DNA fragment described in the
 CC method of the invention.

XX
 SQ Sequence 2866 BP; 593 A; 752 C; 790 G; 731 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 2866;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 ttgcgctcgcgc 17
 |||
 DB 2641 ttgcgctcgcgc 2654

RESULT 20
 AAZ36323/C
 ID AAZ36323 standard; CDNA; 3720 BP.

XX
 AC AAZ36323;

XX
 DT 22-FEB-2000 (first entry)

XX
 DE Mechanical stress induced CDNA encoding protein 405.

XX
 KW Mechanical stress; gene therapy; protein 405; osteoporosis; bone density;
 KW bone development; ss.

XX
 OS Rattus sp.

XX
 FH Key Location/Qualifiers
 FT CDS 45..3177
 FT /tag=a
 FT /product="protein 405"

XX
 EN W09960164-A1.

XX
 PD 25-NOV-1999.

XX
 PF 14-MAY-1999; 99WO-US11066.

XX
 PR 15-MAY-1998; 98US-0085673.

XX
 PA (QUAR-) QUARK BIOTECH INC.

XX
 PI Elinat P, Mor O, Skalter R, Feinstein E, Faerman A;

XX
 DR WPI: 2000-053304/04.

XX
 DR P-PSDB; AAY35669.

XX
 PT Identification of stress induced genes for determining risk and
 PT preventing, treating or controlling osteoporosis

XX
 PS Claim 25; Fig 9A-M; 308pp; English.

XX
 CC The present sequence encodes protein 405, which was identified
 CC using the method of the invention after subjecting rat osteoblasts to
 CC mechanical stress. Expression of the 405 gene was found to be
 CC downregulated in response to mechanical stimulation and stress.

XX
 CC The specification describes a method for the identification of genes
 CC responsive to a specific mechanical stress. The method comprises applying
 CC the mechanical stress to an organism (tissue or cells comprising bone
 CC cells), isolating the specific cellular fractions and extracting mRNA
 CC from them, and differentially analysing the mRNA in comparison with
 CC control samples. The method is used to identify genes whose expression
 CC is responsive to a specific stress. The identified genes are employed in

CC determining risk associated with a physiological or disease state. The
 CC risk determination methods are used for testing a medication for gene
 CC therapy. These medications, or genes identified by the method of the
 CC invention, are used for treating, preventing or controlling a
 CC physiological or disease state (especially osteoporosis or bone density
 CC or other factors causing or contributing to osteoporosis or its symptoms
 CC or other conditions involved in mechanical stress or its lack. The
 CC methods can also be used for advancing research or studies in bone
 CC development.

XX
 SQ Sequence 3720 BP; 946 A; 991 C; 955 G; 828 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 3720;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 gatcctcgaagt 21
 |||
 DB 574 GATCCTCGTCAGG 561

RESULT 21
 AAS07382/C
 ID AAS07382 standard; CDNA; 4027 BP.

XX
 AC AAS07382;

XX
 DT 26-SEP-2001 (first entry)

XX
 DE Human DNA associated with CLASP-5 #2.

XX
 KW Human; CLASP-5; cadherin-like asymmetry protein; immune gateway;
 KW immunogen; antibody; autoimmune disease; rheumatoid arthritis;
 KW multiple sclerosis; leukemia; insulin dependent diabetes mellitus;
 KW acquired immunodeficiency syndrome; AIDS; ss.

XX
 OS Homo sapiens.

XX
 FH Key Location/Qualifiers
 FT CDS 1..2970
 FT /tag=a
 FT /product="protein as displayed in AAU04026"

XX
 FT /partial
 FT /note="No start codon"

XX
 PN W0200142296-A2.

XX
 PD 14-JUN-2001.

XX
 PF 13-DEC-2000; 2000WO-US34163.

XX
 PR 13-DEC-1999; 99US-0170453.

XX
 PR 14-JAN-2000; 2000US-0176195.

XX
 PR 14-FEB-2000; 2000US-0182296.

XX
 PR 11-APR-2000; 2000US-0196267.

XX
 PR 11-APR-2000; 2000US-0196460.

XX
 PR 11-APR-2000; 2000US-0196527.

XX
 PR 11-APR-2000; 2000US-0196528.

XX
 PR 11-APR-2000; 2000US-0547276.

XX
 PR 13-OCT-2000; 2000US-0240503.

XX
 PR 13-OCT-2000; 2000US-0240508.

XX
 PR 13-OCT-2000; 2000US-0240543.

XX
 PR 13-OCT-2000; 2000US-0240539.

XX
 PA (ARBO-) ARBOR VITA CORP.

XX
 PI Lu P, Garman JD, Candia AF;

XX
 DR WPI: 2001-367865/38.

XX
 DR P-PSDB; AAU04026.

XX
 PT CLASP-5 polynucleotides, proteins and antibodies are used to prevent or

PT treat a CLASP-5 mediated disease, such as an autoimmune disease e.g.
 PT Rheumatoid arthritis -
 PS
 XX Disclousure: Fig 9C: 188pp; English.
 CC The sequence encodes a protein presented as AAU04026, included in the
 CC specification which relates to a cadherin-like asymmetry protein,
 CC CLASP-5, which is a transmembrane protein of the immune system involved
 CC in the formation of the immune gateway. CLASP-5, polynucleotides encoding
 CC it and an anti-CLASP-5 antibody are used to prevent or treat a CLASP-5
 CC mediated disease, such as an autoimmune disease caused or exacerbated by
 CC increased activity of TH1 cells. These diseases may include Addison's
 CC disease, rheumatoid arthritis, dermatitis, Multiple Sclerosis, Systemic
 CC Lupus Erythematosus and Autoimmune Thyroiditis. Inflammatory conditions
 CC (e.g. Ischaemia-reperfusion) and responses, Leukaemia, acquired
 CC immunodeficiency syndrome, AIDS, syphilis, typhoid, Burkitt's lymphoma,
 CC Grave's disease and insulin dependent diabetes mellitus. CLASP-5 is used
 CC for inhibiting an immune response in a cell (T cell or B cell) by
 CC interfering with the expression of a CLASP-5 gene in the cell, the
 CC ability of a CLASP-5 protein to bind to another cell or the ability of a
 CC CLASP-5 protein to bind to another protein. The antibody binding CLASP-5
 CC is used to inhibit an immune response in a subject. The polynucleotide is
 CC used to detect CLASP-5 expression in cells and for diagnosis of
 CC diseases and disorders associated with aberrant expression of CLASP-5.
 CC Note: The present sequence is included in the specification but is
 CC not mentioned anywhere in the specification.
 CC
 SQ Sequence 4027 BP; 1162 A; 898 C; 957 G; 1010 T; 0 other;
 XX
 XX
 Query Match 56.0%; Score 14; DB 22; Length 4027;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 9 atccctgcagtg 22
 |||||||||
 Db 1815 ATCCTGTCAGGTG 1802
 RESULT 22
 AAS07373/c
 ID AAS07373 standard; cDNA: 7215 BP.
 XX
 AC AAS07373;
 DT 26-SEP-2001 (first entry)
 XX
 DE Human cDNA encoding CLASP-5.
 XX
 KM Human: CLASP-5; cadherin-like asymmetry protein; immune gateway;
 KM immunogen: antibody; autoimmune disease; rheumatoid arthritis;
 KM multiple sclerosis; leukaemia; insulin dependent diabetes mellitus;
 KM acquired immunodeficiency syndrome; AIDS; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 112..6048
 FT /*tag= a
 FT /product= "CLASP-5"
 FT primer_bind 1498..1519
 FT /*tag= b
 FT /note= "PCR primer HCSgs1"
 FT primer_bind 3087..3113
 FT /*tag= c
 FT /note= "Antisense oligonucleotide AAS07384"
 FT primer_bind 3087..3113
 FT /*tag= d
 FT /note= "Antisense oligonucleotide AAS07384"
 FT primer_bind 3232..3259
 FT /*tag= e
 FT /note= "Antisense oligonucleotide AAS07384"
 FT primer_bind complement (3642..3660)
 FT

FT /*tag= f
 FT /note= "PCR primer HCSAS10b"
 FT primer_bind 4728..4756
 FT /*tag= g
 FT /note= "Antisense oligonucleotide AAS07384"
 XX
 PN WO200142296-A2.
 XX
 PD 14-JUN-2001.
 XX
 XX 13-DEC-2000; 2000MO-US34163.
 PR 13-DEC-1999; 99US-0170453.
 PR 14-JAN-2000; 2000US-0176195.
 PR 14-FEB-2000; 2000US-0182296.
 PR 11-APR-2000; 2000US-0196267.
 PR 11-APR-2000; 2000US-0196460.
 PR 11-APR-2000; 2000US-0196527.
 PR 11-APR-2000; 2000US-0196528.
 PR 11-APR-2000; 2000US-0547276.
 PR 13-OCT-2000; 2000US-0240503.
 PR 13-OCT-2000; 2000US-0240508.
 PR 13-OCT-2000; 2000US-0240543.
 PR 13-OCT-2000; 2000US-0240539.
 PR
 PA (ARBO-) ARBOR VITA CORP.
 XX
 PL Lu P, Garman JD, Candia AF;
 XX
 DR WPI: 2001-367865/38.
 DR P-PSDB: AAU04024.
 XX
 PT CLASP-5 polynucleotides, proteins and antibodies are used to prevent or
 PT treat a CLASP-5 mediated disease, such as an autoimmune disease e.g.
 PT Rheumatoid arthritis -
 PT
 XX
 XX Claim 1; Fig 6; 188pp; English.
 XX
 CC The sequence encodes a cadherin-like asymmetry protein, CLASP-5, which
 CC is a transmembrane protein of the immune system involved in the formation
 CC of the immune gateway. CLASP-5, polynucleotides encoding it and an
 CC anti-CLASP-5 antibody are used to prevent or treat a CLASP-5 mediated
 CC disease, such as an autoimmune disease caused or exacerbated by
 CC increased activity of TH1 cells. These diseases may include Addison's
 CC disease, rheumatoid arthritis, dermatitis, Multiple Sclerosis, Systemic
 CC Lupus Erythematosus and Autoimmune Thyroiditis. Inflammatory conditions
 CC (e.g. Ischaemia-reperfusion) and responses, Leukaemia, acquired
 CC immunodeficiency syndrome, AIDS, syphilis, typhoid, Burkitt's lymphoma,
 CC Grave's disease and insulin dependent diabetes mellitus. CLASP-5 is used
 CC for inhibiting an immune response in a cell (T cell or B cell) by
 CC interfering with the expression of a CLASP-5 gene in the cell, the
 CC ability of a CLASP-5 protein to bind to another cell or the ability of a
 CC CLASP-5 protein to bind to another protein. The antibody binding CLASP-5
 CC is used to inhibit an immune response in a subject. The polynucleotide is
 CC used to detect CLASP-5 expression in cells and for diagnosis of
 CC diseases and disorders associated with aberrant expression of CLASP-5.
 CC
 SQ Sequence 7215 BP; 2005 A; 1734 C; 1721 G; 1755 T; 0 other;
 XX
 Query Match 56.0%; Score 14; DB 22; Length 7215;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 9 atccctgcagtg 22
 |||||||||
 Db 5004 ATCCTGTCAGGTG 4991
 RESULT 23
 AAS07380/c
 ID AAS07380 standard; DNA: 66685 BP.
 XX

FT	/*lag= as
FT	/number= 22
FT	3953. 5808
FT	/*lag= at
FT	/number= 22
XX	
PN	W0200142296-A2.
XX	
PD	14-JUN-2001.
XX	
PF	13-DEC-2000; 2000WO-US34163.
XX	
PR	13-DEC-1999; 99US-0170453.
PR	14-JAN-2000; 2000US-0176195.
PR	14-FEB-2000; 2000US-0182296.
PR	11-APR-2000; 2000US-0196267.
PR	11-APR-2000; 2000US-0196460.
PR	11-APR-2000; 2000US-0196527.
PR	11-APR-2000; 2000US-0196528.
PR	11-APR-2000; 2000US-0547276.
PR	13-OCT-2000; 2000US-0240503.
PR	13-OCT-2000; 2000US-0240508.
PR	13-OCT-2000; 2000US-0240543.
XX	
PA	(ARBO-) ARBOR VITA CORP.
XX	
PI	Lu P, Garman JD, Candia AF;
XX	
XX	WPI; 2001-367865/38.
XX	
PT	CLASP-5 polynucleotides, proteins and antibodies are used to prevent or
PT	treat a CLASP-5 mediated disease, such as an autoimmune disease e.g.
PT	Rheumatoid arthritis -
XX	
PS	Example 2: Fig 7B; 188pp; English.
XX	
CC	The sequence represents a genomic clone containing part of the gene
CC	which encodes cadherin-like asymmetry protein, CLASP-5,
CC	a transmembrane protein of the immune system involved in the formation
CC	of the immune gateway. CLASP-5, polynucleotides encoding it and an
CC	anti-CLASP-5 antibody are used to prevent or treat a CLASP-5 mediated
CC	disease, such as an autoimmune disease caused or exacerbated by
CC	increased activity of TH1 cells. These diseases may include Addison's
CC	disease, Rheumatoid arthritis, dermatitis, Multiple Sclerosis, Systemic
CC	lupus Erythematosus and Autoimmune Thyroiditis, inflammatory conditions
CC	(e.g Ischemia-reperfusion) and responses, leukaemia, acquired
CC	immunodeficiency syndrome, AIDS, syphilis, typhoid, Burkitt's lymphoma,
CC	Grave's disease and insulin dependent diabetes mellitus. CLASP-5 is used
CC	for inhibiting an immune response in a cell (T cell or B cell) by
CC	interfering with the expression of a CLASP-5 gene in the cell, the
CC	ability of a CLASP-5 protein to bind to another cell or the ability of a
Query Match 56.0%; Score 14; DB 22; Length 66685;	
Best Local Similarity 100.0%; Pred. No. 22;	
Matches 14; conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	9 atccctcgtcagtg 22
DB	38189 ATCTCTGTCAGCTG 38176
RESULT 24	
ID	AAC81914 standard; DNA; 273254 BP.
XX	
AC	AAC81914;
XX	
DT	27-FEB-2001 (first entry)
XX	
DE	Chlamydia pneumoniae genome DNA.
XX	
KW	Genome; diagnosis; vaccine; ds.

OS	Chlamydia pneumoniae.
XX	
PN	W020002794-A2.
PD	
XX	18-MAY-2000.
PF	12-NOV-1999; 99WO-US26923.
XX	
PR	12-APR-1998; 98US-0108279.
PR	08-APR-1999; 99US-0128606.
PA	(REGC) UNIV CALIFORNIA.
XX	
PI	Stephens R, Mitchell W, Kalman S, Davis R;
DR	WPI: 2000-376516/32.
XX	
PT	Isolated nucleic acid for use in diagnostic and analytical methods
XX	encodes genomic sequence of Chlamydia pneumoniae -
PS	Claim 2; Page 128-320; 320pp; English.
XX	
CC	This invention describes a novel nucleic acid (N1) encoding a Chlamydia
CC	pneumoniae protein (P1), given in the specification. The isolated nucleic
CC	acid is useful for diagnostic and analytical methods, such as,
CC	hybridization-based assays or amplification-based assays. The protein may
CC	be used for diagnostic purposes, for their enzymatic or structural
CC	activity, or as a vaccine. The invention also describes (1) a probe
CC	comprising a hybridizing fragment of N1; (2) an isolated nucleic acid
CC	(N2) that hybridizes under stringent conditions to N1; (3) an expression
CC	cassette comprising N1 under the transcriptional regulation of a
CC	transcriptional initiation region functional in an expression host, and a
CC	transcriptional termination region; (4) a cell comprising an expression
CC	cassette of (3) as part of an extrachromosomal element or integrated into
CC	the genome of a host cell as a result of induction of the expression
CC	cassette into the host cell, and the cellular progeny of the host cell;
CC	(5) a method for producing a P1 comprising growing a cell of (4) where
CC	the protein is expressed and isolating the protein free of other
CC	proteins; (6) a purified polypeptide composition comprising at least 50
CC	weight % of P1; and (7) a monoclonal antibody binding specifically to the
CC	peptide of (6).
XX	
SO	Sequence 273254 BP; 76423 A; 51054 C; 61965 G; 83812 T; 0 other:
Query Match	56.0%; Score 14; DB 21; Length 273254;
Best Local Similarity	100.0%; Pred. No. 20;
Matches 14; Conservative	0; Mismatches 0; Indels 0; Gaps 0.
OY	8 gatcctgcgtcaagt 21
Db	32246 gatccctgcgtcagt 32259
RESULT 25	
AAH68527	
ID	AAH68527 standard; DNA; 349980 BP.
XX	
AC	AAH68527;
XX	
DT	26-SEP-2001 (first entry)
XX	
DE	C glutamicum coding sequence fragment SEQ ID NO: 7062.
XX	
KM	Corynebacterium; amino acid synthesis; vitamin; saccharide;
KW	organic acid synthesis; ds.
XX	
OS	Corynebacterium glutamicum.
XX	
PN	EPII08790-A2.
XX	
PD	20-JUN-2001.

KW Corynebacterium bacterium; amino acid synthesis; vitamin; saccharide;
 XX organic acid synthesis; ds.
 OS Corynebacterium glutamicum.
 XX
 PN EPI108790-A2.
 XX
 XX 20-JUN-2001.
 PD
 XX 18-DEC-2000; 2000EP-0127688.
 PF
 XX 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOWA) KYOWA HAKKO KOGYO KK.
 XX
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 DR WPI: 2001-376931/40.
 DR P-PSDB: AAG90714.
 PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 XX
 PS Claim 8: SEQ ID NO: 968; 246pp + Sequence listing; English.
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of Corynebacterium bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Corynebacterium bacterium, and identifying a homologue of a gene derived
 CC from Corynebacterium bacterium. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 303 BP; 68 A; 90 C; 87 G; 58 T; 0 other;

Query Match 52.0%; Score 13; DB 22; Length 303;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 cgaatcctcgtcag 19
 |||||||||
 DB 122 CGATCCTCGTCAG 110

RESULT 28
 AAC08252/c
 ID AAC08252 standard; cDNA; 351 BP.
 XX
 AC AAC08252;
 XX
 XX 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST, SEQ ID NO: 12327.
 XX
 KM Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KM gene therapy; chromosome mapping; ss.
 XX
 OS Homo sapiens.
 OS
 PN EPI033401-A2.
 XX

PD 06-SEP-2000.
 XX
 PF 21-FEB-2000; 2000EP-0200610.
 XX
 PR 26-FEB-1999; 99US-0122487.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Mline Edwards J, Duclert A, Giordano J;
 XX
 DR WPI: 2000-500381/45.
 XX
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 XX Claim 1; SEQ ID 12327; 71pp + CD-ROM; English.
 PS
 XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or poly(A) RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 XX
 SQ Sequence 351 BP; 58 A; 105 C; 122 G; 66 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 351;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tctctcgtcagtg 22
 |||||||||
 DB 220 TCTCTCGTCAGGTG 208

RESULT 29
 AAC56641/c
 ID AAC56641 standard; DNA; 353 BP.
 XX
 AC AAC56641;
 XX
 DT 25-JAN-2001 (first entry)
 XX
 XX Eucalyptus grandis transcription factor DNA sequence #512.
 DE
 DE Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
 KW poplar; sweetgum; teak; mahogany; bzIP; G-box binding factor;
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
 KW homeodomain zipper; LIM domain; AP2; ERBs; zinc finger domain;
 KW type 2 Cys2His2; CCAAT box element; MYB; ss.
 XX
 OS Eucalyptus grandis.
 OS
 PN WO200053724-A2.
 XX
 PD 14-SEP-2000.
 PD
 PF 09-MAR-2000; 2000WO-US06112.
 XX
 PR 11-MAR-1999; 99US-0266513.
 PR 18-AUG-1999; 99US-0149485.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.

Novel nucleic acids and peptides derived from open reading frame X,
 useful for treating e.g. cancers, proliferative disorders,
 neurodegenerative disorders and cardiovascular disease -

Claim 5: Page 1368-1369; 5507pp: English.

AACT4446 to AACT7606 encode the proteins given in AAB40237 to AAB43397,
 which represent the human ORF open reading frames 1 to 3161. The ORF
 sequences have activities such as: cytostatic; hepatotropic; villinolytic;
 antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;
 osteoprotective; anticonvulsant; antiallergic; immunosuppressive;
 immunostimulant; cardiast; thrombolytic; coagulant; vasodilator;
 antidiabetic; hypotensive; dermatologic; immunosuppressive;
 antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 antihypertensive; and antianemic. The sequences can be used for determining
 the presence of or predisposition to, or preventing or treating
 pathological conditions associated with an ORF-associated disorder. The
 nucleic acids can be used to express ORF proteins in gene therapy
 vectors. The proteins and nucleic acids may be used to treat cancers,
 proliferative disorders, neurodegenerative disorders, osteoarthritis,
 graft vs host disease, cardiovascular disease, diabetes mellitus,
 hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 allergies, aplastic anemia, burns, wounds, bone and cartilage damage,
 nocturnal haemoglobinuria, antinflammatory disease; to enhance
 coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 470 BP; 93 A; 166 C; 131 G; 77 T; 3 other;

Query Match 52.0%; Score 13; DB 21; Length 470;

Best Local Similarity 100.0%; Pred. No. 1.1e+02; Mismatches 0; Gaps 0;

Matches 13; Conservative 0; Indels 0; Gaps 0;

6 tcgagctcgcga 18

|||||

68 TCGATCTCGCTCA 56

RESULT 32

AAH65984

ID AAH65984 standard; DNA: 480 BP.

AAH65984;

26-SEP-2001 (first entry)

C glutamicum coding sequence fragment SEQ ID NO: 1019.

Corynebacterium glutamicum; amino acid synthesis; vitamin; saccharide;

organic acid synthesis; ds.

Corynebacterium glutamicum.

EP1108790-A2.

20-JUN-2001.

18-DEC-2000; 2000EP-0127688.

16-DEC-1999; 99JP-0377484.

07-APR-2000; 2000JP-0159162.

03-AUG-2000; 2000JP-0280988.

(KYOW) KYOWA HAKKO KOGYO KK.

Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

WPI; 2001-376931/40.

P-SDB; AAG90765.

Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 mutation point of a gene, measuring expression of a gene, analysing
 expression profile or pattern of a gene and identifying homologous gene

Claim 8; SEQ ID NO: 1019; 246pp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein
 sequences from the Corynebacterium glutamicum. These
 are useful for identifying the mutation point of a gene derived from a
 mutant of corynebacterium, measuring expression amount and
 analysing the expression profile or expression pattern of a gene derived
 from Corynebacterium bacteria, and identifying a homologue of a gene derived
 from corynebacterium bacteria. Corynebacterium bacteria are useful for producing
 amino acids, nucleic acids, vitamins, saccharides and organic acids,
 particularly L-lysine. The present sequence is a nucleic acid described
 in the exemplification of the invention.
 Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from the
 European Patent Office.

Sequence 480 BP; 113 A; 101 C; 151 G; 115 T; 0 other;

Query Match 52.0%; Score 13; DB 22; Length 480;

Best Local Similarity 100.0%; Pred. No. 1.1e+02; Mismatches 0; Gaps 0;

Matches 13; Conservative 0; Indels 0; Gaps 0;

13 tcgtcaggtgcag 25

|||||

399 tcgtcaggtgcag 411

RESULT 33

AAAC54077

ID AAAC54077 standard; DNA: 591 BP.

AAAC54077;

18-OCT-2000 (first entry)

Arabidopsis thaliana DNA fragment SEQ ID NO: 76608.

Hybridisation assay; genetic mapping; gene expression control;

protein identification; signal transduction pathway;

metabolic pathway; promoter; termination sequence; ss.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

05-MAR-1999; 99US-0123180.

09-MAR-1999; 99US-0123548.

23-MAR-1999; 99US-0125788.

25-MAR-1999; 99US-0126264.

29-MAR-1999; 99US-0126785.

01-APR-1999; 99US-0127462.

06-APR-1999; 99US-0128234.

08-APR-1999; 99US-0128714.

16-APR-1999; 99US-0129845.

19-APR-1999; 99US-0130077.

21-APR-1999; 99US-0130449.

23-APR-1999; 99US-0130510.

28-APR-1999; 99US-0130891.

30-APR-1999; 99US-0131449.

30-APR-1999; 99US-0132048.

04-MAY-1999; 99US-0132407.

04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139765.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.


```
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match          52.0%; Score 13; DB 21; Length 591;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 attcgatccctcg 15
   |||||
Db 144 attcgatccctcg 156

RESULT 34
AAC35502
ID AAC35502 standard; DNA; 606 BP.
XX
AC AAC35502;
XX
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 10429.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW Protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
PN
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
```

```
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
```

```

PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0156569.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159684.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160778.
PR 21-OCT-1999; 99US-0160779.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-01611404.

```

```

PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 606;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 attcgatccctcg 15
Db 159 attcgatccctcg 171

```

```

RESULT 35
AAC75197
ID AAC75197 standard; cDNA; 623 BP.
AC AAC75197;
XX 08-FEB-2001 (first entry)
DE Human ORFX ORF752 polynucleotide sequence SEQ ID NO:1503.
XX
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX vulnereary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX immunostimulant; osteopathic; antiarthritic; immunosuppressant; cardiant;
XX hypotensive; dermatological; immunosuppressive; antidiabetic;
XX antiviral; antibacterial; antifungal; antirheumatic; antihypoid;
XX antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; cartilage damage; antiinflammatory disease; coagulation;
XX thrombosis; contraceptive; ss.
XX
XX Homo sapiens.
XX
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WC-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach M;
XX
XX MPI: 2000-602362/57.
XX P-PDB: AAB40988.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 5; Page 1245; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnereary;

```


PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159688.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.

PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161044.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161820.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 996;
 Best Local Similarity 100.0%; Pred.No.1.le+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ttctgacctcgt 16
 |||||
 Db 714 TTTCGATCCTCGT 702

RESULT 37
 AAF90596/c
 ID AAF90596 standard; cDNA to mRNA; 1081 BP.
 XX AAF90596;
 AC AAF90596;
 XX
 DT 22-AUG-2001 (first entry)
 XX
 DE Cotton transcription factor GMYB 7 cDNA.
 XX
 KW GMYB 7; R2R3-MYB; transcription factor; cotton;
 KW transgenic plant; ss.
 XX
 OS Gossypium hirsutum cv. Acala SJ-2.
 XX
 FH Key Location/Qualifiers
 FT CDS 86..994
 FT /*tag= a
 XX
 PN WC200140250-A1.
 XX
 PD 07-JUN-2001.
 XX
 PF 01-DEC-2000; 2000WO-US32731.
 XX
 PR 02-DEC-1999; 99US-0453387.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Wilkins TA;
 XX
 DR WPI: 2001-397938/42.
 DR P-PSDB; AAB82483.
 XX
 PT Modulating transcription in plants for improving cotton fiber quality;
 PT comprises introducing a recombinant expression cassette comprising a
 PT promoter operably linked to a sequence encoding a MYB polypeptide -
 XX
 PS Claim 21; Page 43; 50pp; English.
 XX
 CC The present sequence is that of cDNA encoding cotton putative MYB
 CC transcription factor GMYB 7 (see AAB82483). The cDNA was isolated
 CC from a Gossypium hirsutum cv. Acala SJ-2 10-day pre-anthesis fibre
 CC cDNA library. A claimed method of modulating transcription in a
 CC plant involves introducing into the plant a recombinant expression
 CC cassette comprising a promoter operably linked to a heterologous
 CC polynucleotide encoding a MYB protein, especially GMYB 1 or
 CC GMYB 6. The plant is especially cotton, and the promoter directs
 CC expression in cotton fibres or roots. Valuable phenotypes can

CC be conferred on the plant by this method, e.g. increased fibre
CC quality, yield, length, strength or fineness, alteration of root
CC architecture, and enhanced growth.

XX Sequence 1081 BP; 353 A; 228 C; 240 G; 260 T; 0 other;

Query Match 52.0%; Score 13; DB 22; Length 1081;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tctctgcaggtc 22
|||||
Db 309 tctctgcaggtc 297

RESULT 38
AAC41953/C
ID AAC41953 standard; DNA; 1114 BP.

XX AAC41953;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 33748.

XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126785.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.

```

PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155658.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match          52.0%; Score 13; DB 21; Length 1114;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ttgcatactcgt 16
   |||||
Db 6/4 TTTCGATCCTCCT 662

RESULT 39
AAT98707/c
ID AAT98707 standard; DNA: 1284 BP.
XX
AC AAT98707;
XX
DT 09-NOV-1998 (first entry)
XX
DE DNA encoding a S. pneumoniae protein of unknown function.
XX
KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;
KW Immunological response; inoculation; antibody production; inhibitor;
KW T cell immune response; antimicrobial compound; bacterial adhesion;
KW extracellular matrix protein; protein-mediated cell invasion; wound;
KW pathogenesis; ss.
XX
OS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 310..537
FT /*tag= a
XX
PD WO9743303-A1.
XX
PE 20-NOV-1997.
XX
PF 14-MAY-1997; 97WO-US07950.
XX
PR 14-MAY-1996; 96US-0017670.
XX
PA (SMK ) SMITHKLINE BEECHAM CORP.
PI (SMK ) SMITHKLINE BEECHAM PLC.
XX
PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
PI Stodola RK;
XX
DR WPI: 1998-008793/01.
DR P-PSDB: AAM38663.
XX
PT Novel Streptococcus pneumoniae proteins and related DNA - useful for
PT diagnosing anti-microbial agents for treatment of bacterial
PT infections
XX
PS Claim 4: Pages 210-211; 483pp: English.
XX
CC This sequence encodes a Streptococcus pneumoniae protein of unknown
CC function, and represents a DNA sequence of the invention.
CC The DNA sequences were isolated from Streptococcus pneumoniae strain
CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
CC invention can be used to identify compounds which interact with and
CC inhibit or activate the activity of the proteins. Antagonists can be
CC used to treat diseases caused by S. pneumoniae proteins, through genetic
CC immunisation. They can also be used to induce an immunological response
CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
CC of the encoding nucleic acids in a vector adequate to produce antibody
CC and/or T cell immune responses to protect the animal from disease. The
CC proteins can also be used to identify antimicrobial compounds which are
CC capable of inhibiting their bioactivity. In particular the proteins of
CC the invention can be used to prevent adhesion of bacteria to mammalian
CC extracellular matrix proteins on in-dwelling devices or in wounds, to
CC block protein-mediated mammalian cell invasion, and to block the normal
CC progression of pathogenesis in infections initiated other than by the
CC implantation of in-dwelling devices or other surgical techniques.
XX

```

SO Sequence 1284 BP; 382 A; 257 C; 320 G; 324 T; 1 other;

Query Match 52.0%; Score 13; DB 19; Length 1284;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tggatcctcgta 18
|||||

DB 995 TCGATCCTCGTCA 983

RESULT 40

AAV42945/C

ID AAV42945 standard; DNA; 1284 BP.

XX AAV42945;

DT 09-NOV-1998 (first entry)

DE Streptococcus pneumoniae polypeptide coding region.

XX Polypeptide: ORF; open reading frame; infection; bacterial;

KW streptococcal; bacteremia; diagnosis; prophylaxis; ds.

XX Streptococcus pneumoniae.

OS Key Location/Qualifiers

FT CDS 714..1049

XX MO9823631-A1.

XX 04-JUN-1998.

PD 24-NOV-1997; 97WO-US21976.

XX 27-NOV-1996; 96US-0031879.

PA (SMK) SMITHKLINE BEECHAM CORP.

XX (SMK) SMITHKLINE BEECHAM PLC.

PI Black MT, Hodson JE, Knowles DJC, Lonetto MA, Nicholas RO;

DR Reid RH, Zarfos PN;

XX WPI; 1998-322654/28.

DR P-SDB; AAW62663.

XX Streptococcus pneumoniae polynucleotides - useful for developing

PT products for diagnosis, prevention and treatment of infections e.g.

XX pneumonia, bacteremia, meningitis or endocarditis

PS Claim 1; Page 35-36; 181pp; English.

XX The sequence is that of a Streptococcal polypeptide coding region.

CC The polypeptide can potentially be used for the diagnosis and

CC prevention of bacterial infections, especially SP infection.

CC It may be used for the treatment of diseases such as otitis media,

CC conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural

CC empyema, endocarditis or infection of the cerebrospinal fluid.

XX Sequence 1284 BP; 382 A; 257 C; 320 G; 324 T; 1 other;

Query Match 52.0%; Score 13; DB 19; Length 1284;

Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tggatcctcgta 18
|||||

DB 995 TCGATCCTCGTCA 983

RESULT 41

ID AAC36484

XX AAC36484 standard; DNA; 1359 BP.

AC AAC36484;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 13980.

XX Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

XX metabolic pathway; promoter; termination sequence; ss.

CS Arabidopsis thaliana.

XX EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 29-MAR-1999; 99US-0126264.

PR 01-APR-1999; 99US-0126785.

PR 06-APR-1999; 99US-0127462.

PR 08-APR-1999; 99US-0128234.

PR 16-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0129845.

PR 21-APR-1999; 99US-0130077.

PR 23-APR-1999; 99US-0130449.

PR 28-APR-1999; 99US-0130510.

PR 30-APR-1999; 99US-0130891.

PR 04-MAY-1999; 99US-0131449.

PR 05-MAY-1999; 99US-0132048.

PR 06-MAY-1999; 99US-0132484.

PR 07-MAY-1999; 99US-0132485.

PR 11-MAY-1999; 99US-0132486.

PR 14-MAY-1999; 99US-0132487.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 18-MAY-1999; 99US-0134221.

PR 19-MAY-1999; 99US-0134370.

PR 20-MAY-1999; 99US-0134768.

PR 21-MAY-1999; 99US-0134941.

PR 24-MAY-1999; 99US-0135124.

PR 25-MAY-1999; 99US-0135353.

PR 27-MAY-1999; 99US-0135629.

PR 28-MAY-1999; 99US-0136021.

PR 03-JUN-1999; 99US-0136392.

PR 07-JUN-1999; 99US-0137222.

PR 08-JUN-1999; 99US-0137528.

PR 10-JUN-1999; 99US-0137724.

PR 10-JUN-1999; 99US-0138094.

PR 14-JUN-1999; 99US-0138540.

PR 16-JUN-1999; 99US-0138847.

PR 17-JUN-1999; 99US-0139119.

PR 18-JUN-1999; 99US-0139452.

PR 18-JUN-1999; 99US-0139453.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR	18-JUN-1999;	99US-0139459;	PR	20-AUG-1999;	99US-0149929;
PR	18-JUN-1999;	99US-0139460;	PR	23-AUG-1999;	99US-0149902;
PR	18-JUN-1999;	99US-0139461;	PR	23-AUG-1999;	99US-0149930;
PR	18-JUN-1999;	99US-0139463;	PR	25-AUG-1999;	99US-0150566;
PR	18-JUN-1999;	99US-0139465;	PR	26-AUG-1999;	99US-0150884;
PR	18-JUN-1999;	99US-0139750;	PR	27-AUG-1999;	99US-0151065;
PR	21-JUN-1999;	99US-0139817;	PR	27-AUG-1999;	99US-0151066;
PR	22-JUN-1999;	99US-0139899;	PR	30-AUG-1999;	99US-0151080;
PR	23-JUN-1999;	99US-0140353;	PR	31-AUG-1999;	99US-0151303;
PR	24-JUN-1999;	99US-0140354;	PR	01-SEP-1999;	99US-0151438;
PR	28-JUN-1999;	99US-0140695;	PR	07-SEP-1999;	99US-0151930;
PR	29-JUN-1999;	99US-0140823;	PR	10-SEP-1999;	99US-0152363;
PR	30-JUN-1999;	99US-0140991;	PR	13-SEP-1999;	99US-0153070;
PR	01-JUL-1999;	99US-0141287;	PR	15-SEP-1999;	99US-0153758;
PR	01-JUL-1999;	99US-0141842;	PR	16-SEP-1999;	99US-0154039;
PR	02-JUL-1999;	99US-0142055;	PR	20-SEP-1999;	99US-0154779;
PR	06-JUL-1999;	99US-0142390;	PR	22-SEP-1999;	99US-0155139;
PR	08-JUL-1999;	99US-0142803;	PR	23-SEP-1999;	99US-0155486;
PR	09-JUL-1999;	99US-0142920;	PR	24-SEP-1999;	99US-0155659;
PR	12-JUL-1999;	99US-0142977;	PR	28-SEP-1999;	99US-0156458;
PR	13-JUL-1999;	99US-0143542;	PR	29-SEP-1999;	99US-0156596;
PR	14-JUL-1999;	99US-0143624;	PR	04-OCT-1999;	99US-0157117;
PR	15-JUL-1999;	99US-0144005;	PR	05-OCT-1999;	99US-0157753;
PR	16-JUL-1999;	99US-0144085;	PR	06-OCT-1999;	99US-0157865;
PR	16-JUL-1999;	99US-0144086;	PR	07-OCT-1999;	99US-0158029;
PR	19-JUL-1999;	99US-0144325;	PR	08-OCT-1999;	99US-0158232;
PR	19-JUL-1999;	99US-0144331;	PR	12-OCT-1999;	99US-0158369;
PR	19-JUL-1999;	99US-0144332;	PR	13-OCT-1999;	99US-0159293;
PR	19-JUL-1999;	99US-0144333;	PR	13-OCT-1999;	99US-0159294;
PR	19-JUL-1999;	99US-0144334;	PR	14-OCT-1999;	99US-0159329;
PR	19-JUL-1999;	99US-0144335;	PR	14-OCT-1999;	99US-0159380;
PR	20-JUL-1999;	99US-0144352;	PR	14-OCT-1999;	99US-0159381;
PR	20-JUL-1999;	99US-0144632;	PR	14-OCT-1999;	99US-0159637;
PR	21-JUL-1999;	99US-0144884;	PR	18-OCT-1999;	99US-0159638;
PR	21-JUL-1999;	99US-0144884;	PR	21-OCT-1999;	99US-0159584;
PR	21-JUL-1999;	99US-0145086;	PR	21-OCT-1999;	99US-0160741;
PR	21-JUL-1999;	99US-0145088;	PR	21-OCT-1999;	99US-0160767;
PR	22-JUL-1999;	99US-0145085;	PR	21-OCT-1999;	99US-0160768;
PR	22-JUL-1999;	99US-0145087;	PR	21-OCT-1999;	99US-0160770;
PR	22-JUL-1999;	99US-0145089;	PR	21-OCT-1999;	99US-0160814;
PR	22-JUL-1999;	99US-0145192;	PR	21-OCT-1999;	99US-0160815;
PR	23-JUL-1999;	99US-0145145;	PR	22-OCT-1999;	99US-0160980;
PR	23-JUL-1999;	99US-0145218;	PR	22-OCT-1999;	99US-0160981;
PR	26-JUL-1999;	99US-0145224;	PR	22-OCT-1999;	99US-0160989;
PR	26-JUL-1999;	99US-0145276;	PR	25-OCT-1999;	99US-0161404;
PR	27-JUL-1999;	99US-0145913;	PR	25-OCT-1999;	99US-0161405;
PR	27-JUL-1999;	99US-0145918;	PR	25-OCT-1999;	99US-0161406;
PR	27-JUL-1999;	99US-0145919;	PR	26-OCT-1999;	99US-0161359;
PR	28-JUL-1999;	99US-0145991;	PR	26-OCT-1999;	99US-0161360;
PR	02-AUG-1999;	99US-0146386;	PR	26-OCT-1999;	99US-0161361;
PR	02-AUG-1999;	99US-0146388;	PR	28-OCT-1999;	99US-0161920;
PR	03-AUG-1999;	99US-0146389;	PR	28-OCT-1999;	99US-0161982;
PR	04-AUG-1999;	99US-0147204;	PR	28-OCT-1999;	99US-0161993;
PR	04-AUG-1999;	99US-0147302;	PR	29-OCT-1999;	99US-01

XX C glutamicum coding sequence fragment SEQ ID NO: 2952.
XX
XX Corynebacterium: amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
XX Corynebacterium glutamicum.
OS
XX EPI108790-A2.
PN
XX 20-JUN-2001.
PD
XX
XX 18-DEC-2000; 2000EP-0127688.
PF
XX
XX 16-DEC-1999; 99JP-0377484.
PR
XX 07-APR-2000; 2000JP-0159162.
PR
XX 03-AUG-2000; 2000JP-0280988.
PR
XX (KYOW) KYOWA HAKKO KOGYO KK.
PA
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI: 2001-376931/40.
DR P-PSDB; AAG92698.
XX
XX Novel polynucleotides derived from Corynebacterium, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PS
XX
XX Claim 8; SEQ ID NO: 2952; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from corynebacterium, and identifying a homologue of a gene derived
CC from corynebacterium. Corynebacterium are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX
SQ Sequence 1443 BP; 241 A; 382 C; 380 G; 440 T; 0 other;

OS Arabidopsis thaliana.
XX
XX EPI033405-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR
XX 05-MAR-1999; 99US-0123180.
PR
XX 09-MAR-1999; 99US-0123548.
PR
XX 23-MAR-1999; 99US-0125788.
PR
XX 25-MAR-1999; 99US-0126264.
PR
XX 29-MAR-1999; 99US-0126785.
PR
XX 01-APR-1999; 99US-0127462.
PR
XX 06-APR-1999; 99US-0128234.
PR
XX 08-APR-1999; 99US-0128714.
PR
XX 16-APR-1999; 99US-0129845.
PR
XX 19-APR-1999; 99US-0130077.
PR
XX 21-APR-1999; 99US-0130449.
PR
XX 23-APR-1999; 99US-0130510.
PR
XX 28-APR-1999; 99US-0130891.
PR
XX 28-APR-1999; 99US-0131449.
PR
XX 30-APR-1999; 99US-0132407.
PR
XX 30-APR-1999; 99US-0132408.
PR
XX 04-MAY-1999; 99US-0132484.
PR
XX 05-MAY-1999; 99US-0132485.
PR
XX 06-MAY-1999; 99US-0132486.
PR
XX 06-MAY-1999; 99US-0132487.
PR
XX 07-MAY-1999; 99US-0132487.
PR
XX 11-MAY-1999; 99US-0134256.
PR
XX 14-MAY-1999; 99US-0134218.
PR
XX 14-MAY-1999; 99US-0134219.
PR
XX 14-MAY-1999; 99US-0134221.
PR
XX 14-MAY-1999; 99US-0134370.
PR
XX 18-MAY-1999; 99US-0134768.
PR
XX 19-MAY-1999; 99US-0134941.
PR
XX 20-MAY-1999; 99US-0135124.
PR
XX 20-MAY-1999; 99US-0135353.
PR
XX 24-MAY-1999; 99US-0135629.
PR
XX 25-MAY-1999; 99US-0136021.
PR
XX 27-MAY-1999; 99US-0136392.
PR
XX 28-MAY-1999; 99US-0136782.
PR
XX 01-JUN-1999; 99US-0137222.
PR
XX 03-JUN-1999; 99US-0137528.
PR
XX 04-JUN-1999; 99US-0137502.
PR
XX 07-JUN-1999; 99US-0137724.
PR
XX 08-JUN-1999; 99US-0138094.
PR
XX 10-JUN-1999; 99US-0138540.
PR
XX 10-JUN-1999; 99US-0138847.
PR
XX 14-JUN-1999; 99US-0139119.
PR
XX 16-JUN-1999; 99US-0139452.
PR
XX 16-JUN-1999; 99US-0139453.
PR
XX 17-JUN-1999; 99US-0139492.
PR
XX 18-JUN-1999; 99US-0139454.
PR
XX 18-JUN-1999; 99US-0139455.
PR
XX 18-JUN-1999; 99US-0139456.
PR
XX 18-JUN-1999; 99US-0139457.
PR
XX 18-JUN-1999; 99US-0139458.
PR
XX 18-JUN-1999; 99US-0139459.
PR
XX 18-JUN-1999; 99US-0139460.
PR
XX 18-JUN-1999; 99US-0139461.
PR
XX 18-JUN-1999; 99US-0139462.
PR
XX 18-JUN-1999; 99US-0139463.
PR
XX 18-JUN-1999; 99US-0139750.
PR
XX 18-JUN-1999; 99US-0139763.
PR
XX 21-JUN-1999; 99US-0139817.
PR
XX 22-JUN-1999; 99US-0139899.
PR
XX 23-JUN-1999; 99US-0140353.
PR
XX 23-JUN-1999; 99US-0140354.
PR
XX 24-JUN-1999; 99US-0140695.
PR
XX 28-JUN-1999; 99US-0140823.
PR
XX 29-JUN-1999; 99US-0140991.

```

PR 30-JUN-1999; 99US-0141827.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145226.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.

```

```

PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159285.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 1497;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tctcgcagtg 22
    |||||
Db 174 tctcgcagtg 186

RESULT 44
ID AAF70984 standard; DNA: 1566 BP.
XX
AC AAF70984;
XX
DE 30-APR-2001 (first entry)
XX
C. glutamicum SRT protein nucleotide sequence SPO ID NO:1.
XX
KW Corynebacterium glutamicum; stress; resistance; tolerance; SRT;
KW fine chemical production; organic acid; proteinogetic amino acid;
KW nonproteinogetic amino acid; purine base; pyrimidine base; nucleoside;
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;
KW evolutionary study; environmental hazard; fermentation; ds.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100804-A2.
XX

```

PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-1B00922.
XX
PR 25-JUN-1999; 99US-0141031.
PR 01-JUL-1999; 99DE-1030429.
PR 01-JUL-1999; 99US-0142692.
PR 08-JUL-1999; 99DE-1031413.
PR 08-JUL-1999; 99DE-1031457.
PR 08-JUL-1999; 99DE-1031541.
PR 09-JUL-1999; 99DE-1032209.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032914.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99US-0151214.
PR 31-AUG-1999; 99DE-1041382.
XX
RA (BADI) BASF AG.
XX
PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G, Lee H;
PI Kim H;
XX
DR WPI: 2001-061972/07.
DR P-PSDB; AAB78871.
XX
PT New isolated Corynebacterium glutamicum nucleic acid encoding a stress,
PT tolerance or resistance protein, for production or modulation of
PT production of fine chemicals, such as, e.g. amino acids, lipids,
PT carbohydrates, or enzymes -
XX
PS Claim 3; Page 111-113; 526pp; English.
XX
XX AAF70984 to AAF71133 encode the Corynebacterium glutamicum stress,
CC resistance, and tolerance (SRT) proteins given in AAB78871 to AAB79020.
CC The C. glutamicum SRT genes (I) can be used in vectors (II) for
CC expression in host cells and production of fine chemicals, such as, an
CC organic acid, a proteohogenic or nonproteohogenic amino acid (preferred),
CC a purine or pyrimidine base, a nucleoside, a nucleotide, a lipid, a
CC saturated or unsaturated fatty acid, a diol, a carbohydrate, an aromatic
CC compound, a vitamin, a cofactor, a polyketide, or an enzyme. The fine
CC chemical production can be modulated. The presence of (I) or the SRT
CC proteins (III) encoded by them are used for diagnosing the presence
CC or activity of Corynebacterium diphtheriae, (I), (II), (III) and host
CC cells containing them can be used to map the genomes of organisms related
CC to C. glutamicum, to identify and localise C. glutamicum sequences of
CC interest, in evolutionary studies, in determination of SRT protein
CC regions required for function, in modulating the SRT protein activity,
CC and in modulating the activity of an SRT pathway. (II) are used to permit
CC C. glutamicum to survive in an environment that is normally
CC environmentally or chemically hazardous to it. (I) and protein molecules
CC encoded by it increase the survival of C. glutamicum to chemical and
CC environmental hazards and provide a means for continued growth and
CC multiplication in large scale fermentative growth conditions. By
CC increasing the growth rate or maintaining a normal growth rate in poor or
CC toxic conditions, the yield, production and/or efficiency or production
CC of fine chemicals from a culture may be increased.
XX
XX Sequence 1566 BP; 275 A; 406 C; 403 G; 482 T; 0 other;

AC AAC38498;
XX
DT 17-OCT-2000 (First entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 21191.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135123.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.

Query Match 52.0%; Score 13; DB 22; Length 1566;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 4 ttgcattcctgt 16
Db 171 ttgcattcctgt 183

RESULT 45
AAC38498
ID AAC38498 standard; DNA: 1782 BP.
XX

```
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.

PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153707.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160818.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 1782;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tctctgcagtg 22
      |||||
Db 299 tctctgcagtg 311
```

Search completed: December 26, 2001, 14:17:06
Job time: 7018 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:55:33 ; Search time 143.1 Seconds
(without alignments)
39.566 Million cell updates/sec

Title: US-09-396-196f-8

Perfect score: 25

Sequence: 1 gcatttcgacccctcgtcagtgacg 25

Scoring table: OLIGO_NUC
Gapext 60.0 , Gapext 60.0

Searched: 351203 seqs, 11323899 residues

Word size : 9

Total number of hits satisfying chosen parameters: 9549

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/PCNUS.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCNUS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	14	56.0	66	3	US-09-023-228B-96
6	14	56.0	2104	1	US-08-682-193A-1
7	12	48.0	19	1	US-07-936-421-11
8	12	48.0	50	2	US-08-850-049-122
9	12	48.0	50	2	US-08-050-478-122
10	12	48.0	50	4	US-09-414-117-122
11	12	48.0	61	4	US-08-687-421-385
12	12	48.0	500	4	US-09-248-528-1
13	12	48.0	500	4	US-09-549-108-1
14	12	48.0	500	4	US-09-549-111-1
15	12	48.0	500	4	US-09-549-106-1
16	12	48.0	500	4	US-09-550-394-1
17	12	48.0	615	2	US-08-465-485A-22
18	12	48.0	615	3	US-09-080-285-22
19	12	48.0	623	6	5506344-3
20	12	48.0	717	2	US-08-465-485A-20
21	12	48.0	717	3	US-09-080-285-20
22	12	48.0	760	1	US-08-405-702A-11
23	12	48.0	831	3	US-08-927-128-3
24	12	48.0	831	6	5459251-3
25	12	48.0	831	6	5506344-4
26	12	48.0	864	2	US-08-595-868C-11
27	12	48.0	867	4	US-09-139-819A-11

28	12	48.0	876	1	US-08-726-725-5	Sequence 5, Appl
29	12	48.0	911	4	US-09-126-109-5	Sequence 5, Appl
30	12	48.0	911	5	PCT-US93-06251-3	Sequence 3, Appl
31	12	48.0	1058	4	US-08-818-112-45	Sequence 45, Appl
32	12	48.0	1411	1	US-07-952-87-10	Sequence 10, Appl
33	12	48.0	1411	6	5210025-3	Patent No. 5210025
34	12	48.0	1516	4	US-09-248-528-3	Sequence 3, Appl
35	12	48.0	1516	4	US-09-549-111-3	Sequence 3, Appl
36	12	48.0	1516	4	US-09-549-111-3	Sequence 3, Appl
37	12	48.0	1516	4	US-09-549-106-3	Sequence 3, Appl
38	12	48.0	1516	4	US-09-550-394-3	Sequence 3, Appl
39	12	48.0	1517	3	US-09-248-528-2	Sequence 2, Appl
40	12	48.0	1517	4	US-09-549-108-2	Sequence 2, Appl
41	12	48.0	1517	4	US-09-549-111-2	Sequence 2, Appl
42	12	48.0	1517	4	US-09-549-106-2	Sequence 2, Appl
43	12	48.0	1517	4	US-09-550-394-2	Sequence 2, Appl
44	12	48.0	1602	4	US-08-987-151-4	Sequence 4, Appl
45	12	48.0	1785	4	US-08-987-151-1	Sequence 4, Appl
46	12	48.0	1785	4	US-08-987-151-3	Sequence 3, Appl
47	12	48.0	1846	2	US-08-365-486A-16	Sequence 16, Appl
48	12	48.0	1846	4	US-08-880-342-16	Sequence 16, Appl
49	12	48.0	1889	2	US-09-026-587-2	Sequence 2, Appl
50	12	48.0	1889	2	US-09-227-420-2	Sequence 2, Appl
51	12	48.0	2115	1	US-08-285-641-19	Sequence 19, Appl
52	12	48.0	2230	1	US-08-217-327-5	Sequence 5, Appl
53	12	48.0	2854	1	US-08-121-713D-57	Sequence 57, Appl
54	12	48.0	2854	1	US-08-835-268-57	Sequence 57, Appl
55	12	48.0	2854	2	US-09-060-692-57	Sequence 57, Appl
56	12	48.0	2854	2	US-08-833-391-57	Sequence 57, Appl
57	12	48.0	2854	5	PCT-US94-10151A-57	Sequence 57, Appl
58	12	48.0	3000	1	US-07-841-997A-3	Sequence 3, Appl
59	12	48.0	3000	1	US-08-290-301-3	Sequence 3, Appl
60	12	48.0	3505	1	US-08-454-097-34	Sequence 34, Appl
61	12	48.0	3505	1	US-08-468-036-43	Sequence 43, Appl
62	12	48.0	3505	2	US-08-376-843-43	Sequence 43, Appl
63	12	48.0	3505	4	US-08-185-359-34	Sequence 34, Appl
64	12	48.0	4127	3	US-09-487-368A-10	Sequence 10, Appl
65	12	48.0	4265	4	US-09-061-709-1	Sequence 1, Appl
66	12	48.0	4825	6	5459251-1	Patent No. 5459251
67	12	48.0	5086	2	US-08-465-485A-19	Sequence 19, Appl
68	12	48.0	5086	2	US-08-365-486A-14	Sequence 14, Appl
69	12	48.0	5086	3	US-09-080-285-19	Sequence 19, Appl
70	12	48.0	5086	4	US-08-880-342-14	Sequence 14, Appl
71	12	48.0	5086	5	PCT-US93-05651-4	Sequence 4, Appl
72	12	48.0	5086	6	PCT-US93-06251-2	Sequence 2, Appl
73	12	48.0	5104	6	5506344-1	Patent No. 5506344
74	12	48.0	6085	4	US-09-029-603-4	Sequence 4, Appl
75	12	48.0	4403765	4	US-09-103-840A-2	Sequence 2, Appl
76	12	48.0	4403765	4	US-09-103-840A-2	Sequence 2, Appl
77	12	48.0	4411529	4	US-09-103-840A-1	Sequence 1, Appl
78	12	48.0	4411529	4	US-09-103-840A-1	Sequence 1, Appl
79	11	44.0	19	3	US-08-504-538A-16	Sequence 16, Appl
80	11	44.0	19	5	PCT-US95-09307-16	Sequence 16, Appl
81	11	44.0	22	3	US-08-963-927-3	Sequence 3, Appl
82	11	44.0	22	4	US-09-481-810-3	Sequence 3, Appl
83	11	44.0	23	1	US-08-343-785-15	Sequence 15, Appl
84	11	44.0	23	2	US-08-462-221-15	Sequence 15, Appl
85	11	44.0	23	3	US-08-462-221-15	Sequence 15, Appl
86	11	44.0	23	2	US-08-946-458-13	Sequence 13, Appl
87	11	44.0	24	2	US-08-771-624B-2	Sequence 2, Appl
88	11	44.0	25	1	US-08-343-785-14	Sequence 14, Appl
89	11	44.0	25	2	US-08-462-221-14	Sequence 14, Appl
90	11	44.0	25	3	US-08-946-458-13	Sequence 13, Appl
91	11	44.0	27	2	US-08-462-221-13	Sequence 13, Appl
92	11	44.0	27	3	US-08-946-458-13	Sequence 13, Appl
93	11	44.0	30	1	US-08-479-487-64	Sequence 64, Appl
94	11	44.0	37	1	US-08-723-896-10	Sequence 10, Appl
95	11	44.0	47	2	US-08-771-624B-16	Sequence 16, Appl
96	11	44.0	55	1	US-08-681-935-17	Sequence 17, Appl
97	11	44.0	55	2	US-08-771-624B-3	Sequence 3, Appl
98	11	44.0	55	4	US-08-939-323-17	Sequence 17, Appl
99	11	44.0	61	3	US-09-023-228B-47	Sequence 47, Appl
100	11	44.0	61	3	US-09-023-228B-59	Sequence 59, Appl

ALIGNMENTS

```
RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIRA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "Biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 9.9e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gcatcgatcctcgtcagtgacg 25
    |||||||||||||||||||
Db 99 GCATTGATCCTCCTGTCAGTGACG 123

RESULT 2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
```

```
ADDRESS: No. 5869719artis Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meligs, J. Timothy
; REGISTRATION NUMBER: 38,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "Biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 9.9e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gcatcgatcctcgtcagtgacg 25
    |||||||||||||||||||
Db 99 GCATTGATCCTCCTGTCAGTGACG 123

RESULT 3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
```

```

APPLICATION NUMBER: US/08/411.768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pB030A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=117
OTHER INFORMATION: /product="Biotin synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="biob"
OTHER INFORMATION: /number=1
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=2295
OTHER INFORMATION: /function="involved in pimeoyl-CoA synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /gene="bioc"
OTHER INFORMATION: /number=3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=3750
OTHER INFORMATION: /ec_number=2.6.1.62
OTHER INFORMATION: /product="DAPA synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="bioA"
OTHER INFORMATION: /number=5
OTHER INFORMATION: /standard_name=
OTHER INFORMATION: "S-Adenosyl-L-methionine:8-aminoo-7-oxononanoate
OTHER INFORMATION: aminoitransf."
FEATURE:
NAME/KEY: CDS
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start=5098
OTHER INFORMATION: /function="unknown, involved in biotin synthesis"
OTHER INFORMATION: /product="protein"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="ORF1"
OTHER INFORMATION: /number=6
FEATURE:
NAME/KEY: -10-signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="promoter plac"
FEATURE:
NAME/KEY: -35-signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name="promoter plac"

```

```

FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name="bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name="bioA RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name="ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name="rho-independent
OTHER INFORMATION: transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function="promoter plac"
OTHER INFORMATION: /evidence=EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

```

```

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 gcattcgatcctcgtcagtgacg 25
    |||||||1111111111111111
DB 215 GCATTGCAATCCTCGTCAGGTCAG 239

```

```

RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect
; SOFTWARE: Version 5.1

```



```

      ZIP: 02019
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Diskette
      OPERATING SYSTEM: DOS
      SOFTWARE: FastSeq Version 1.5
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/682.193A
      FILING DATE: 17-JUL-1996
      CLASSIFICATION: 435
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: 7-181730
      FILING DATE: 18-JUL-1995
      ATTORNEY/AGENT INFORMATION:
      NAME: DAVID, RESNICK S
      REGISTRATION NUMBER: 34,235
      REFERENCE/DOCKET NUMBER: 46643
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: 617-523-3400
      TELEFAX: 617-523-6440
      TELEX: 200291 STRE
      INFORMATION FOR SEQ ID NO: 1:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 2104 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: linear
      MOLECULE TYPE: Genomic DNA
      HYPOTHETICAL: NO
      ANTI-SENSE: NO
      FRAGMENT TYPE:
      ORIGINAL SOURCE:
      ORGANISM: Brevibacterium flavum
      STRAIN: MJ-233
      FEATURE:
      NAME/KEY: Coding Sequence
      LOCATION: 556..1855
      OTHER INFORMATION:
      US-08-682-193A-1

Query Match      56.0%; Score 14; DB 1; Length 2104;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 ttccgataccgcgc 17
    |||||||||
Db 1635 TTTCGATCCTCGTC 1648

RESULT 7
US-07-936-421-11/c
; Sequence 11, Application US/07936421
; Patent No. 5750390
; GENERAL INFORMATION:
; APPLICANT: James D. Thompson
; APPLICANT: Kenneth G. Draper
; TITLE OF INVENTION: METHOD AND REAGENT FOR
; TITLE OF INVENTION: TREATMENT OF DISEASES CAUSED
; TITLE OF INVENTION: BY EXPRESSION OF THE BCL-2
; TITLE OF INVENTION: GENE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS (Version 5.0)

```

```

      SOFTWARE: WordPerfect (Version 5.1)
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/07/936,421
      FILING DATE: 19920826
      CLASSIFICATION: 435
      PRIOR APPLICATION DATA:
      PRIOR APPLICATION NUMBER:
      PRIOR APPLICATION DATA: described below:
      APPLICATION NUMBER:
      FILING DATE:
      ATTORNEY/AGENT INFORMATION:
      NAME: Warburg, Richard J.
      REGISTRATION NUMBER: 32,327
      REFERENCE/DOCKET NUMBER: 197/243
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (213) 489-1600
      TELEFAX: (213) 955-0440
      TELEX: 67-3510
      INFORMATION FOR SEQ ID NO: 11:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 19
      TYPE: NUCLEIC ACID
      STRANDEDNESS: single
      TOPOLOGY: linear
      US-07-936-421-11

Query Match      48.0%; Score 12; DB 1; Length 19;
Best Local Similarity 100.0%; Pred. No. 2,1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 14 cgtcagatcgacg 25
    |||||||||
Db 18 CGTGAGTGCAG 7

RESULT 8
US-08-850-049-122/c
; Sequence 12, Application US/08850049
; Patent No. 5965726
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT:
; TITLE OF INVENTION: METHOD OF ELIMINATING
; TITLE OF INVENTION: INHIBITORY/INSTABILITY REGIONS OF MRNA
; NUMBER OF SEQUENCES: 130
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/850,049
; FILING DATE: 02-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/050,478
; FILING DATE: 26-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/02908
; FILING DATE: 29-MAR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/858,747

```

FILED DATE: 27-MAR-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MORRY, MARY J.
REGISTRATION NUMBER: 34,398
REFERENCE/DOCKET NUMBER: 2026-4006US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)758-4800
TELEFAX: (212)751-6849
INFORMATION FOR SEQ ID NO: 122:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
US-08-850-049-122

Query Match 48.0%; Score 12; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 cgaatcctgcga 18
Db 43 cgaatcctgcga 32

RESULT 9
US-08-050-478-122/c
Sequence 122, Application US/08050478
Patent No. 5972596
GENERAL INFORMATION:
APPLICANT:
APPLICANT:
TITLE OF INVENTION: METHOD OF ELIMINATING
TITLE OF INVENTION: INHIBITORY/INSTABILITY REGIONS OF mRNA
NUMBER OF SEQUENCES: 130
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/050,478
FILING DATE: 26-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/02908
FILING DATE: 29-MAR-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/858,747
FILING DATE: 27-MAR-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MORRY, MARY J.
REGISTRATION NUMBER: 34,398
REFERENCE/DOCKET NUMBER: 2026-4006US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)758-4800
TELEFAX: (212)751-6849
INFORMATION FOR SEQ ID NO: 122:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 BASE PAIRS
TYPE: NUCLEIC ACID

STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
US-08-050-478-122

Query Match 48.0%; Score 12; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 cgaatcctgcga 18
Db 43 cgaatcctgcga 32

RESULT 10
US-09-414-117-122/c
Sequence 122, Application US/09414117
Patent No. 6291664
GENERAL INFORMATION:
APPLICANT:
APPLICANT:
TITLE OF INVENTION: METHOD OF ELIMINATING
TITLE OF INVENTION: INHIBITORY/INSTABILITY REGIONS OF mRNA
NUMBER OF SEQUENCES: 130
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/414,117
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/850,049
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/02908
FILING DATE: 29-MAR-1993
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/858,747
FILING DATE: 27-MAR-1992
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MORRY, MARY J.
REGISTRATION NUMBER: 34,398
REFERENCE/DOCKET NUMBER: 2026-4006US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)758-4800
TELEFAX: (212)751-6849
INFORMATION FOR SEQ ID NO: 122:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
US-09-414-117-122

Query Match 48.0%; Score 12; DB 4; Length 50;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 cgaatcctgcagca 18
|||||
Db 43 CGATCTCTGTCGA 32

RESULT 11
US-08-687-421-385
Sequence 385, Application US/08687421
Patent No. 6177557
GENERAL INFORMATION:
APPLICANT: Gold, Larry
APPLICANT: Janjic, Nebojsa
APPLICANT: Tasseil, Diane
TITLE OF INVENTION: HIGH-AFFINITY LIGANDS OF BASIC
TITLE OF INVENTION: FIBROBLAST GROWTH FACTOR AND
TITLE OF INVENTION: THROMBIN
NUMBER OF SEQUENCES: 445
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB storage
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,421
FILING DATE: 08-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/195,005
FILING DATE: 10-FEBRUARY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE: 22-APRIL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/219,012
FILING DATE: 28-MARCH-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,333
FILING DATE: 11-NOVEMBER-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/714,131
FILING DATE: 10-JUNE-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/536,428
FILING DATE: 11-JUNE-1990
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX07/PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 385:
SEQUENCE CHARACTERISTICS:
LENGTH: 61 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-687-421-385

Query Match 48.0%; Score 12; DB 4; Length 61;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 8 gatcctgcagc 19
|||||

Db 37 GATCTCTGTCAG 48

RESULT 12
US-09-248-528-1/c
Sequence 1, Application US/09248528C
Patent No. 6153415
GENERAL INFORMATION:
APPLICANT: Orieli, Patrick J
APPLICANT: Padmakumar, Rugmini
APPLICANT: Kim, Sang H
TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile
TITLE OF INVENTION: Hydratase from a Thermophilic Bacillus
FILE REFERENCE: MSU 4.1-401
CURRENT APPLICATION NUMBER: US/09/248,528C
CURRENT FILING DATE: 1999-02-10
EARLIER APPLICATION NUMBER: 60/083,485
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 500
TYPE: DNA
ORGANISM: BR449
FEATURE:
NAME/KEY: rRNA
LOCATION: (1)..(500)
US-09-248-528-1

Query Match 48.0%; Score 12; DB 3; Length 500;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tctctgcagcgt 21
|||||
Db 484 TCCTCTGTCAGCT 473

RESULT 13
US-09-549-108-1/c
Sequence 1, Application US/09549108
Patent No. 6214603
GENERAL INFORMATION:
APPLICANT: Orieli, Patrick J
APPLICANT: Padmakumar, Rugmini
APPLICANT: Kim, Sang H
TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile
TITLE OF INVENTION: Hydratase from a Thermophilic Bacillus
FILE REFERENCE: MSU 4.1-486
CURRENT APPLICATION NUMBER: US/09/549,108
CURRENT FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 60/083,485
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 09/248,528
PRIOR FILING DATE: 1999-02-10
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 500
TYPE: DNA
ORGANISM: BR449
FEATURE:
NAME/KEY: rRNA
LOCATION: (1)..(500)
US-09-549-108-1

Query Match 48.0%; Score 12; DB 4; Length 500;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 10 tctctgcagcgt 21

Db 484 TCCTCGTCAGGT 473

|||||
RESULT 14
US-09-549-111-1/c
; Sequence 1, Application US/09549111
; Patent No. 6228633
; GENERAL INFORMATION:
; APPLICANT: Oriol, Patrick J
; APPLICANT: Padmakumar, Rugmini
; APPLICANT: Kim, Sang H
; TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile
; FILE REFERENCE: MSU 4.1-489
; CURRENT APPLICATION NUMBER: US/09/549,111
; CURRENT FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/083,485
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 09/248,528
; PRIOR FILING DATE: 1999-02-10
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 500
; TYPE: DNA
; ORGANISM: BR449
; FEATURE:
; NAME/KEY: rRNA
; LOCATION: (1)..(500)
US-09-549-111-1

Query Match 48.0%; Score 12; DB 4; Length 500;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tcctcgtcaggt 21
|||||
Db 484 TCCTCGTCAGGT 473

RESULT 15
US-09-549-106-1/c
; Sequence 1, Application US/09549106
; Patent No. 6242242
; GENERAL INFORMATION:
; APPLICANT: Oriol, Patrick J
; APPLICANT: Padmakumar, Rugmini
; APPLICANT: Kim, Sang H
; TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile
; FILE REFERENCE: MSU 4.1-487
; CURRENT APPLICATION NUMBER: US/09/549,106
; CURRENT FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/083,485
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 09/248,528
; PRIOR FILING DATE: 1999-02-10
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 500
; TYPE: DNA
; ORGANISM: BR449
; FEATURE:
; NAME/KEY: rRNA
; LOCATION: (1)..(500)
US-09-549-106-1

Query Match 48.0%; Score 12; DB 4; Length 500;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 tcctcgtcaggt 21
|||||
Db 484 TCCTCGTCAGGT 473

RESULT 16
US-09-550-394-1/c
; Sequence 1, Application US/09550394
; Patent No. 6287828
; GENERAL INFORMATION:
; APPLICANT: Oriol, Patrick J
; APPLICANT: Padmakumar, Rugmini
; APPLICANT: Kim, Sang H
; TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile
; FILE REFERENCE: MSU 4.1-488
; CURRENT APPLICATION NUMBER: US/09/550,394
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/083,485
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 09/248,528
; PRIOR FILING DATE: 1999-02-10
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 500
; TYPE: DNA
; ORGANISM: BR449
; FEATURE:
; NAME/KEY: rRNA
; LOCATION: (1)..(500)
US-09-550-394-1

Query Match 48.0%; Score 12; DB 4; Length 500;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tcctcgtcaggt 21
|||||
Db 484 TCCTCGTCAGGT 473

RESULT 17
US-08-465-485A-22/c
; Sequence 22, Application US/08465485A
; Patent No. 5831066
; GENERAL INFORMATION:
; APPLICANT: Reed, John
; TITLE OF INVENTION: Regulation of bcl-2 Gene Expression
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBION, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. Jefferson Davis Hwy., Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,485A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/124,256
; FILING DATE: 20-SEP-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/840,716
FILING DATE: 21-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/288,692
FILING DATE: 22-DEC-1988
ATTORNEY/AGENT INFORMATION:
NAME: Fortney, Andrew D.
REGISTRATION NUMBER: 34,600
REFERENCE/DOCKET NUMBER: 3335-070-55 CONT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (408) 436-2070
TELEFAX: (408) 436-2075
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 615 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..615
US-08-465-485A-22

Query Match 48.0%; Score 12; DB 2; Length 615;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 cgtcaggtcag 25
|||||
DB 366 CGTCAGGTGCGAG 355

RESULT 18
US-09-080-285-22/c
Sequence 22, Application US/09080285
Patent No. 6040181
GENERAL INFORMATION:
APPLICANT: Reed, John
TITLE OF INVENTION: Regulation of bcl-2 Gene Expression
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
P.C.
STREET: 1755 S. Jefferson Davis Hwy., Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/080,285
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/465,485
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/124,256
FILING DATE: 20-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/840,716
FILING DATE: 21-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/288,692
FILING DATE: 22-DEC-1988
ATTORNEY/AGENT INFORMATION:

NAME: Fortney, Andrew D.
REGISTRATION NUMBER: 34,600
REFERENCE/DOCKET NUMBER: 3335-070-55 CONT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (408) 436-2070
TELEFAX: (408) 436-2075
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 615 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..615
US-09-080-285-22

Query Match 48.0%; Score 12; DB 3; Length 615;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 cgtcaggtcag 25
|||||
DB 366 CGTCAGGTGCGAG 355

RESULT 19
5506344-3/c
Patent No. 5506344
APPLICANT: TSUJIMOTO, YOSHITAKE, CROCE, CARLO A.
TITLE OF INVENTION: ANTIBODIES SPECIFIC FOR BCL-2 GENE PRODUCT
NUMBER OF SEQUENCES: 5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/435,193
FILING DATE: 05-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 228,704
FILING DATE: 18-APR-1994
APPLICATION NUMBER: 994,941
FILING DATE: 23-DEC-1992
APPLICATION NUMBER: 663,010
FILING DATE: 19-MAR-1991
APPLICATION NUMBER: 883,687
FILING DATE: 09-JUL-1986
SEQ ID NO: 3;
LENGTH: 623
5506344-3

Query Match 48.0%; Score 12; DB 6; Length 623;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 cgtcaggtcag 25
|||||
DB 366 CGTCAGGTGCGAG 355

RESULT 20
US-08-465-485A-20/c
Sequence 20, Application US/08465485A
Patent No. 5831066
GENERAL INFORMATION:
APPLICANT: Reed, John
TITLE OF INVENTION: Regulation of bcl-2 Gene Expression
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
P.C.
STREET: 1755 S. Jefferson Davis Hwy., Suite 400
CITY: Arlington

STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,485A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/124,256
FILING DATE: 20-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/840,716
FILING DATE: 21-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/288,692
FILING DATE: 22-DEC-1988
ATTORNEY/AGENT INFORMATION:
NAME: Fortney, Andrew D.
REGISTRATION NUMBER: 34,600
REFERENCE/DOCKET NUMBER: 3335-070-55 CONT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (408) 436-2070
TELEFAX: (408) 436-2075
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 717 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..717
US-08-465-485A-20

Query Match 48.0%; Score 12; DB 2; Length 717;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 cgtcaggtgcag 25
|||||
DB 366 CGTCAGGTGCAG 355

RESULT 21
US-09-080-285-20/c
Sequence 20, Application US/09080285
Patent No. 6040181
GENERAL INFORMATION:
APPLICANT: Reed, John
TITLE OF INVENTION: Regulation of bcl-2 Gene Expression
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUBADT,
ADDRESSEE: P.C.
STREET: 1755 S. Jefferson Davis Hwy., Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/080,285

FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/465,485
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/124,256
FILING DATE: 20-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/840,716
FILING DATE: 21-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/288,692
FILING DATE: 22-DEC-1988
ATTORNEY/AGENT INFORMATION:
NAME: Fortney, Andrew D.
REGISTRATION NUMBER: 34,600
REFERENCE/DOCKET NUMBER: 3335-070-55 CONT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (408) 436-2070
TELEFAX: (408) 436-2075
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 717 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..717
US-09-080-285-20

Query Match 48.0%; Score 12; DB 3; Length 717;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 cgtcaggtgcag 25
|||||
DB 366 CGTCAGGTGCAG 355

RESULT 22
US-08-405-702A-11/c
Sequence 11, Application US/08405702A
Patent No. 5789389
GENERAL INFORMATION:
APPLICANT: Tarasiewicz, Dariusz G
APPLICANT: Schott, Brigitte
APPLICANT: Holzmayer, Tatiana A.
APPLICANT: Robinson, Igor B
TITLE OF INVENTION: BCL2 Derived Genetic Elements Associated
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Wilcoff, Ltd.
STREET: 10 South Wacker Drive, Suite 3000
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,702A
FILING DATE: 17-MAR-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: No. 5789389nan, Kevin E

REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 95,332
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
TELEX: 910-221-5317
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 760 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 21..740
US-08-405-702A-11

Query Match
Best Local Similarity 100.0%; Score 12; DB 1; Length 760;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 cgtcaggtgcag 25
|||||
Db 386 CGTCAGGTGCAG 375

RESULT 23
US-08-927-128-3
Sequence 3, Application US/08927128
Patent No. 6127150
GENERAL INFORMATION:
APPLICANT: Coolidge, Thomas
APPLICANT: Wagner, Fred
APPLICANT: ven Heeke, Gino
APPLICANT: Schuster, Sheldon
APPLICANT: Stout, Jay
APPLICANT: Wylie, Dwane
TITLE OF INVENTION: PURIFICATION DIRECTED CLOSING OF PEPTIDES
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESS: Merchant & Gould
STREET: 3100 No. 6127150west Center, 90 S. 7th Street
CITY: Minneapolis
STATE: MN
COUNTRY: U.S.A.
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/927,128
FILING DATE: 05-SEP-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/680,004
FILING DATE: 15-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: Carter, Charles G
REGISTRATION NUMBER: 35,093
REFERENCE/DOCKET NUMBER: 8648,20SD1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612/332-5300
TELEFAX: 612/332-9081
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 780 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1..780
OTHER INFORMATION:
US-08-927-128-3

Query Match
Best Local Similarity 100.0%; Score 12; DB 3; Length 780;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 ttccatcctcgt 16
|||||
Db 532 TTCCATCCTCCT 543

RESULT 24
5459251-3/C
Patent No. 5459251
APPLICANT: Tsujimoto, Yoshide;Croce, Carlo A.
TITLE OF INVENTION: DNA MOLECULES HAVING HUMAN BCL-2 GENE
SEQUENCES
NUMBER OF SEQUENCES: 4
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/228,704
FILING DATE: 18-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 994,941
FILING DATE: 23-DEC-1992
APPLICATION NUMBER: 663,010
FILING DATE: 19-MAR-1991
APPLICATION NUMBER: 883,687
FILING DATE: 09-JUL-1986
SEQ ID NO:3
LENGTH: 831
5459251-3

Query Match
Best Local Similarity 100.0%; Score 12; DB 6; Length 831;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 cgtcaggtgcag 25
|||||
Db 512 CGTCAGGTGCAG 501

RESULT 25
5506344-4/C
Patent No. 5506344
APPLICANT: TSUJIMOTO, YOSHIHIDE;CROCE, CARLO A.
TITLE OF INVENTION: ANTIBODIES SPECIFIC FOR BCL-2GENE PRODUCT
NUMBER OF SEQUENCES: 5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/435,193
FILING DATE: 05-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 228,704
FILING DATE: 18-APR-1994
APPLICATION NUMBER: 994,941
FILING DATE: 23-DEC-1992
APPLICATION NUMBER: 663,010
FILING DATE: 19-MAR-1991
APPLICATION NUMBER: 883,687
FILING DATE: 09-JUL-1986
SEQ ID NO:4
LENGTH: 831

5506344-4

Query Match 48.0%; Score 12; DB 6; Length 831;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 cgtcagtgagc 25
|||||
Db 512 CGTCAGTCGAG 501

RESULT 26
US-08-595-868C-11
; Sequence 11, Application US/08595868C
; Patent No. 5962270
; GENERAL INFORMATION:
; APPLICANT: Wagner, Fred
; APPLICANT: Stout, Jay
; APPLICANT: Henriksen, Dennis
; APPLICANT: Partridge, Bruce
; APPLICANT: Holmquist, Bart
; APPLICANT: Frank, Julie
; TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN FRAGMENTS AND US
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 5962270 West Center, 90 S. 7th Street
; CITY: Minneapolis
; STATE: MN
; COUNTRY: U.S.A.
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/595,868C
; FILING DATE: 06-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Carter, Charles G
; REGISTRATION NUMBER: 35,093
; REFERENCE/DOCKET NUMBER: 8648.590501
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612/332-5300
; TELEFAX: 612/332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 864 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1..864
; OTHER INFORMATION:
US-08-595-868C-11

Query Match 48.0%; Score 12; DB 2; Length 864;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 ttgcattcctcgt 16
|||||
Db 532 TTGCATCTCTCGT 543

RESULT 27
US-09-139-819A-11
; Sequence 11, Application US/09139819A
; Patent No. 6251635
; GENERAL INFORMATION:
; APPLICANT: WAGNER, Fred W.
; APPLICANT: STOUT, Jay S.
; APPLICANT: HENRIKSEN, Dennis B.
; APPLICANT: PARTRIDGE, Bruce E.
; APPLICANT: HOLMQUIST, Bart
; APPLICANT: FRANK, Julie A.
; TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN
; TITLE OF INVENTION: FRAGMENTS AND USE THEREOF IN THE PREPARATION OF CALCITONIN
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/139,819A
; FILING DATE: 25-AUG-1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/595,868
; FILING DATE: 06-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 089187/0144
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 867 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..864
US-09-139-819A-11

Query Match 48.0%; Score 12; DB 4; Length 867;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 ttgcattcctcgt 16
|||||
Db 532 TTGCATCTCTCGT 543

RESULT 28
US-08-726-725-5
; Sequence 5, Application US/08726725
; Patent No. 5773290

GENERAL INFORMATION:
APPLICANT: Gould, Michael N.
APPLICANT: Chen, Kai-Shun
TITLE OF INVENTION: MAMMARY GLAND-SPECIFIC PROMOTERS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: U.S.A.
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/726,725
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Baker, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 960296,93863
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5709
TELEFAX: (414) 271-3552
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 876 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-726-725-5

Query Match 48.0%; Score 12; DB 1; Length 876;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 atccctgcagc 20
|||||
Db 310 ATCCTGCTGAGC 321

RESULT 29
US-09-126-109-5/c
Sequence 5, Application US/09126109
Patent No. 6171856
GENERAL INFORMATION:
APPLICANT: Thispen, Anice
APPLICANT: Hohmeier, Hans-Ewald
APPLICANT: Newgard, Christopher B.
APPLICANT: Ungert, Roger H.
APPLICANT: Shimabukuro, Michio
APPLICANT: Chen, Guaxun
APPLICANT: Rhodes, Christopher J.
APPLICANT: Hugl, Sigrun R.
APPLICANT: Cousin, Sharon
TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
TITLE OF INVENTION: TO NO-MEDIATED CYTOTOXICITY
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/126,109
FILING DATE: 30-JUL-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/055,092
FILING DATE: 30-JUL-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US Unknown
FILING DATE: 03-MAR-1998
ATTORNEY/AGENT INFORMATION:
NAME: McMillian, Nabeela R.
REGISTRATION NUMBER: P-43,363
REFERENCE/DOCKET NUMBER: UTSD:560
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 911 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-126-109-5

Query Match 48.0%; Score 12; DB 4; Length 911;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 cgtcaggtgcag 25
|||||
Db 512 GCTCAGCTGAGC 501

RESULT 30
PCT-US93-06251-3/c
Sequence 3, Application PC/US9306251
GENERAL INFORMATION:
APPLICANT: Wickstrom, Eric and Rife, Jason P.
TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/06251
FILING DATE: 19930630
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8586
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:

LENGTH: 911 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US93-06251-3

Query Match 48.0%; Score 12; DB 5; Length 911;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 cgtcagtcag 25
|||||
Db 512 CGTCAGTCAG 501

RESULT 31
US-08-818-112-45/C
Sequence 45, Application US/08818112
Patent No. 6290969
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 1058 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-45

Query Match 48.0%; Score 12; DB 4; Length 1058;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 ttcgattcct 16
|||||
Db 16 TTCGATTCCT 5

RESULT 32
US-07-952-817-10
Sequence 10, Application US/07952817
Patent No. 5356796
GENERAL INFORMATION:
APPLICANT: Keller, John W.
TITLE OF INVENTION: A Repressor Protein and Gene for Regulating
TITLE OF INVENTION: Expression of Polypeptides and Its Use in the Preparation of
TITLE OF INVENTION: 2,2-Dialkylglycine Decarboxylase of Pseudomonas Cepacia
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: US
ZIP: 20005-3315

Query Match 48.0%; Score 12; DB 6; Length 1411;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/952,817
FILING DATE: 19920928
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 01120.0002-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4400
TELEFAX: 202-408-4000
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1411 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-952-817-10

Query Match 48.0%; Score 12; DB 1; Length 1411;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatcgtatcc 12
|||||
Db 655 GCATTCGATCC 666

RESULT 33
US-07-952-817-10
Sequence 10, Application US/07952817
Patent No. 5356796
GENERAL INFORMATION:
APPLICANT: Keller, John W.
TITLE OF INVENTION: REPRESSOR PROTEIN GENE FOR REGULATING
EXPRESSION OF POLYPEPTIDES AND ITS USE IN THE PREPARATION OF
2,2-DIALKYLGLYCINE DECARBOXYLASE OF PSEUDOMONAS CEPACIA
NUMBER OF SEQUENCES: 18
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/501,814
FILING DATE: 30-MAR-1990
SEQ ID NO: 3
LENGTH: 1411
US-07-952-817-10

Query Match 48.0%; Score 12; DB 6; Length 1411;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcattcgatcc 12
|||||
Db 655 gcattcgatcc 666

RESULT 34
US-09-248-3/C
; Sequence 3, Application US/09248528C
; Patent No. 6153415
; GENERAL INFORMATION:
; APPLICANT: Oriol, Patrick J
; APPLICANT: Padmakumar, Rugmini
; TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrite
; FILE REFERENCE: MSU 4.1-401
; CURRENT FILING DATE: 1999-02-10
; EARLIER FILING DATE: 1998-04-29
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1516
; TYPE: DNA
; ORGANISM: Bacillus pallidus
; FEATURE:
; NAME/KEY: rRNA
; LOCATION: (1)..(1516)
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: 226930/GenBank
; DATABASE ENTRY DATE: 1997-05-14
US-09-248-528-3

Query Match 48.0%; Score 12; DB 3; Length 1516;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tcctcgatcagt 21
|||||
Db 484 TCCTCGTCAGGT 473

RESULT 35
US-09-549-108-3/C
; Sequence 3, Application US/09549108
; Patent No. 6214603
; GENERAL INFORMATION:
; APPLICANT: Oriol, Patrick J
; APPLICANT: Padmakumar, Rugmini
; APPLICANT: Kim, Sang H
; TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrite
; FILE REFERENCE: MSU 4.1-486
; CURRENT FILING DATE: 2000-04-13
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083,485
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 09/248,528
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1516
; TYPE: DNA
; ORGANISM: Bacillus pallidus
; FEATURE:
; NAME/KEY: rRNA
; LOCATION: (1)..(1516)
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: 226930/GenBank
; DATABASE ENTRY DATE: 1997-05-14

US-09-549-108-3

Query Match 48.0%; Score 12; DB 4; Length 1516;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tcctcgatcagt 21
|||||
Db 484 TCCTCGTCAGGT 473

RESULT 36
US-09-549-111-3/C
; Sequence 3, Application US/09549111
; Patent No. 6228633
; GENERAL INFORMATION:
; APPLICANT: Oriol, Patrick J
; APPLICANT: Padmakumar, Rugmini
; APPLICANT: Kim, Sang H
; TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrite
; FILE REFERENCE: MSU 4.1-489
; CURRENT FILING DATE: 2000-04-13
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083,485
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 09/248,528
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1516
; TYPE: DNA
; ORGANISM: Bacillus pallidus
; FEATURE:
; NAME/KEY: rRNA
; LOCATION: (1)..(1516)
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: 226930/GenBank
; DATABASE ENTRY DATE: 1997-05-14
US-09-549-111-3

Query Match 48.0%; Score 12; DB 4; Length 1516;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tcctcgatcagt 21
|||||
Db 484 TCCTCGTCAGGT 473

RESULT 37
US-09-549-106-3/C
; Sequence 3, Application US/09549106
; Patent No. 6242242
; GENERAL INFORMATION:
; APPLICANT: Oriol, Patrick J
; APPLICANT: Padmakumar, Rugmini
; APPLICANT: Kim, Sang H
; TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrite
; FILE REFERENCE: MSU 4.1-487
; CURRENT FILING DATE: 2000-04-13
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083,485
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 09/248,528
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3

```
LENGTH: 1516
TYPE: DNA
ORGANISM: Bacillus pallidus
FEATURE:
NAME/KEY: rRNA
LOCATION: (1)..(1516)
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Z26930/GenBank
DATABASE ENTRY DATE: 1997-05-14
US-09-549-106-3
```

```
Query Match
Best Local Similarity 48.0%; Score 12; DB 4; Length 1516;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 10 tcctcgtagt 21
Db 484 TCCTCGTAGT 473
```

```
RESULT 38
US-09-550-394-3/c
Sequence 3, Application US/09550394
Patent No. 6287828
GENERAL INFORMATION:
APPLICANT: Oriel, Patrick J
APPLICANT: Padmakumar, Rugmini
TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile
FILE REFERENCE: MSU 4.1-488
CURRENT APPLICATION NUMBER: US/09/550,394
PRIOR FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/083,485
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 09/248,528
PRIOR FILING DATE: 1999-02-10
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 3
LENGTH: 1516
TYPE: DNA
ORGANISM: Bacillus pallidus
FEATURE:
NAME/KEY: rRNA
LOCATION: (1)..(1516)
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Z26930/GenBank
DATABASE ENTRY DATE: 1997-05-14
US-09-550-394-3
```

```
Query Match
Best Local Similarity 48.0%; Score 12; DB 4; Length 1516;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 10 tcctcgtagt 21
Db 484 TCCTCGTAGT 473
```

```
RESULT 39
US-09-248-528-2/c
Sequence 2, Application US/09248528C
Patent No. 6153415
GENERAL INFORMATION:
APPLICANT: Oriel, Patrick J
APPLICANT: Padmakumar, Rugmini
TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile
FILE REFERENCE: MSU 4.1-401
```

```
CURRENT APPLICATION NUMBER: US/09/248,528C
CURRENT FILING DATE: 1999-02-10
EARLIER APPLICATION NUMBER: 60/083,485
EARLIER FILING DATE: 1998-04-29
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
LENGTH: 1517
TYPE: DNA
ORGANISM: Bacillus sp.
FEATURE:
NAME/KEY: rRNA
LOCATION: (1)..(1517)
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Z26929
DATABASE ENTRY DATE: 1998-07-02
US-09-248-528-2
```

```
Query Match
Best Local Similarity 48.0%; Score 12; DB 3; Length 1517;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 10 tcctcgtagt 21
Db 487 TCCTCGTAGT 476
```

```
RESULT 40
US-09-549-108-2/c
Sequence 2, Application US/09549108
Patent No. 6214603
GENERAL INFORMATION:
APPLICANT: Oriel, Patrick J
APPLICANT: Padmakumar, Rugmini
TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile
FILE REFERENCE: MSU 4.1-486
CURRENT APPLICATION NUMBER: US/09/549,108
CURRENT FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: 60/083,485
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 09/248,528
PRIOR FILING DATE: 1999-02-10
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
LENGTH: 1517
TYPE: DNA
ORGANISM: Bacillus sp.
FEATURE:
NAME/KEY: rRNA
LOCATION: (1)..(1517)
DATABASE ACCESSION NUMBER: Z26929
DATABASE ENTRY DATE: 1998-07-02
US-09-549-108-2
```

```
Query Match
Best Local Similarity 48.0%; Score 12; DB 4; Length 1517;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 10 tcctcgtagt 21
Db 487 TCCTCGTAGT 476
```

```
RESULT 41
US-09-549-111-2/c
Sequence 2, Application US/09549111
Patent No. 6228633
GENERAL INFORMATION:
```

```

: APPLICANT: Oriel, Patrick J
: APPLICANT: Padmakumar, Rugmini
: APPLICANT: Kim, Sang H
: TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile
: FILE REFERENCE: MSU 4.1-489
: CURRENT APPLICATION NUMBER: US/09/549,111
: CURRENT FILING DATE: 2000-04-13
: PRIOR APPLICATION NUMBER: 60/083,485
: PRIOR FILING DATE: 1998-04-29
: PRIOR APPLICATION NUMBER: 09/248,528
: PRIOR FILING DATE: 1999-02-10
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2
: LENGTH: 1517
: TYPE: DNA
: ORGANISM: Bacillus sp.
: NAME/KEY: rRNA
: LOCATION: (1)..(1517)
: DATABASE ACCESSION NUMBER: Z26929
: DATABASE ENTRY DATE: 1998-07-02
: US-09-549-111-2

```

```

Query Match          48.0%; Score 12; DB 4; Length 1517;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 tcctcgtcaggt 21
Db 487 TCCTCGTCAGGT 476

```

```

RESULT 42
US-09-549-106-2/c
: Sequence 2, Application US/09549106
: Patent No. 6242242
: GENERAL INFORMATION:
: APPLICANT: Oriel, Patrick J
: APPLICANT: Padmakumar, Rugmini
: APPLICANT: Kim, Sang H
: TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile
: FILE REFERENCE: MSU 4.1-487
: CURRENT APPLICATION NUMBER: US/09/549,106
: CURRENT FILING DATE: 2000-04-13
: PRIOR APPLICATION NUMBER: 60/083,485
: PRIOR FILING DATE: 1998-04-29
: PRIOR APPLICATION NUMBER: 09/248,528
: PRIOR FILING DATE: 1999-02-10
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2
: LENGTH: 1517
: TYPE: DNA
: ORGANISM: Bacillus sp.
: NAME/KEY: rRNA
: LOCATION: (1)..(1517)
: DATABASE ACCESSION NUMBER: Z26929
: DATABASE ENTRY DATE: 1998-07-02
: US-09-549-106-2

```

```

Query Match          48.0%; Score 12; DB 4; Length 1517;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 tcctcgtcaggt 21
Db 487 TCCTCGTCAGGT 476

```

```

RESULT 43
US-09-550-394-2/c
: Sequence 2, Application US/09550394
: Patent No. 6287828
: GENERAL INFORMATION:
: APPLICANT: Oriel, Patrick J
: APPLICANT: Padmakumar, Rugmini
: APPLICANT: Kim, Sang H
: TITLE OF INVENTION: Method for Producing Amide Compounds Using a Nitrile
: FILE REFERENCE: MSU 4.1-488
: CURRENT APPLICATION NUMBER: US/09/550,394
: CURRENT FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/083,485
: PRIOR FILING DATE: 1998-04-29
: PRIOR APPLICATION NUMBER: 09/248,528
: PRIOR FILING DATE: 1999-02-10
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2
: LENGTH: 1517
: TYPE: DNA
: ORGANISM: Bacillus sp.
: NAME/KEY: rRNA
: LOCATION: (1)..(1517)
: DATABASE ACCESSION NUMBER: Z26929
: DATABASE ENTRY DATE: 1998-07-02
: US-09-550-394-2

```

```

Query Match          48.0%; Score 12; DB 4; Length 1517;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 tcctcgtcaggt 21
Db 487 TCCTCGTCAGGT 476

```

```

RESULT 44
US-08-987-151-4
: Sequence 4, Application US/08987151
: Patent No. 6162617
: GENERAL INFORMATION:
: APPLICANT: Jaskunas Jr., Stanley R.
: APPLICANT: Zhao, Genshi
: APPLICANT: Peery, Robert B.
: APPLICANT: Burgett, Stanley G.
: APPLICANT: Rostock Jr., Paul R.
: TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESS: Eli Lilly and Company
: STREET: Lilly Corporate Center
: CITY: Indianapolis
: STATE: Indiana
: COUNTRY: U.S.
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/987,151
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Webster, Thomas D.

```

```

: REGISTRATION NUMBER: 39,872
: REFERENCE/DOCKET NUMBER: X-11748
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-276-3334
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1602 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: US-08-987-151-4

```

```

Query Match          48.0%; Score 12; DB 4; Length 1602;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 13 tcgtcaggtgca 24
|||||
Db 1398 TCCTCAGGTGCA 1409

```

```

RESULT 45
US-08-987-151-1
: Sequence 1, Application US/08987151
: Patent No. 6162617

```

```

: GENERAL INFORMATION:
: APPLICANT: Jaskunas Jr., Stanley R.

```

```

: APPLICANT: Zhao, Genshi

```

```

: APPLICANT: Peery, Robert B.

```

```

: APPLICANT: Burgett, Stanley G.

```

```

: APPLICANT: Rostek Jr., Paul R.

```

```

: TITLE OF INVENTION: Streptococcus pneumoniae Gene Sequence
: TITLE OF INVENTION: drag

```

```

: NUMBER OF SEQUENCES: 4

```

```

: CORRESPONDENCE ADDRESS:

```

```

: ADDRESSEE: Eli Lilly and Company

```

```

: STREET: Lilly Corporate Center

```

```

: CITY: Indianapolis

```

```

: STATE: Indiana

```

```

: COUNTRY: U.S.

```

```

: ZIP: 46285

```

```

: COMPUTER READABLE FORM:

```

```

: MEDIUM TYPE: Floppy disk

```

```

: COMPUTER: IBM PC compatible

```

```

: OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

: SOFTWARE: Patent Release #1.0, Version #1.30

```

```

: CURRENT APPLICATION DATA:

```

```

: APPLICATION NUMBER: US/08/987,151

```

```

: FILING DATE:

```

```

: CLASSIFICATION:

```

```

: ATTORNEY/AGENT INFORMATION:

```

```

: NAME: Webster, Thomas D.

```

```

: REGISTRATION NUMBER: 39,872

```

```

: REFERENCE/DOCKET NUMBER: X-11748

```

```

: TELECOMMUNICATION INFORMATION:

```

```

: TELEPHONE: 317-276-3334

```

```

: INFORMATION FOR SEQ ID NO: 1:

```

```

: SEQUENCE CHARACTERISTICS:

```

```

: LENGTH: 1785 base pairs

```

```

: TYPE: nucleic acid

```

```

: STRANDEDNESS: single

```

```

: TOPOLOGY: linear

```

```

: MOLECULE TYPE: DNA (genomic)

```

```

: HYPOTHETICAL: NO

```

```

: ANTI-SENSE: NO

```

```

: FEATURE:

```

```

Query Match          48.0%; Score 12; DB 4; Length 1785;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 13 tcgtcaggtgca 24
|||||
Db 1305 TCCTCAGGTGCA 1316

```

```

Search completed: December 26, 2001, 12:55:38
Job time: 6970 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:10:27 : Search time 4619.78 Seconds
(Without alignments)
58.151 Million cell updates/sec

Title: US-09-396-196f-8

Perfect score: 25

Sequence: 1 gatttcgattcctcgtcgaagtgatgacg 25

Scoring table: OLIGO-MUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 9

Total number of hits satisfying chosen parameters: 373098

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthm:*
3: em_estlin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_est1:*
11: gb_est2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_pro:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16	64.0	300	10	A0113702
2	16	64.0	300	10	A0113702
3	16	64.0	300	10	A0116641
4	16	64.0	300	10	A0116641
5	16	64.0	300	10	A0116641
6	16	64.0	300	10	A0116641
7	16	64.0	300	10	A0116641
8	16	64.0	300	10	A0116641
9	16	64.0	300	10	A0116641
10	16	64.0	300	10	A0116641
11	16	64.0	300	10	A0116641
12	16	64.0	300	10	A0116641

13	C	13	60.0	1096	11	BF144032
14	C	14	60.0	1126	11	BE795069
15	C	15	60.0	1195	11	W06020
16	C	16	60.0	196	10	A1210255
17	C	17	60.0	208	11	BF713183
18	C	18	60.0	257	11	BC408765
19	C	19	60.0	268	11	AV537540
20	C	20	60.0	286	11	BC382504
21	C	21	60.0	291	11	BF994483
22	C	22	60.0	304	11	BF402559
23	C	23	60.0	305	11	C72075
24	C	24	60.0	327	11	A1441334
25	C	25	60.0	360	11	C39749
26	C	26	60.0	368	10	AM699577
27	C	27	60.0	378	10	AL370675
28	C	28	60.0	381	13	FR0035439
29	C	29	60.0	420	10	BE021456
30	C	30	60.0	421	10	A1938935
31	C	31	60.0	429	13	AQ205063
32	C	32	60.0	430	10	A1272538
33	C	33	60.0	452	13	AQ129442
34	C	34	60.0	454	12	AK017623
35	C	35	60.0	457	11	BC726684
36	C	36	60.0	473	13	AO815087
37	C	37	60.0	486	10	A0066422
38	C	38	60.0	494	10	A1507864
39	C	39	60.0	518	11	BF007983
40	C	40	60.0	521	10	AL379335
41	C	41	60.0	524	11	BF634261
42	C	42	60.0	532	13	CNS04KAN
43	C	43	60.0	548	10	AL381376
44	C	44	60.0	559	13	AO528128
45	C	45	60.0	579	11	BI424083
46	C	46	60.0	586	11	BC080788
47	C	47	60.0	610	13	AO655910
48	C	48	60.0	615	11	BC646892
49	C	49	60.0	642	10	AM953979
50	C	50	60.0	644	11	BC067650
51	C	51	60.0	652	11	BC453113
52	C	52	60.0	658	10	AM993297
53	C	53	60.0	659	11	BE821306
54	C	54	60.0	672	10	BF781869
55	C	55	60.0	678	13	AQ361397
56	C	56	60.0	698	10	BE273024
57	C	57	60.0	738	11	BI161170
58	C	58	60.0	742	10	A1226887
59	C	59	60.0	746	10	AM959798
60	C	60	60.0	774	11	BC122704
61	C	61	60.0	775	11	BC855614
62	C	62	60.0	792	13	BC081016
63	C	63	60.0	797	13	CNS06V4X
64	C	64	60.0	812	11	BF867013
65	C	65	60.0	812	11	BI197672
66	C	66	60.0	861	13	CNS03BCG
67	C	67	60.0	881	13	CNS04574
68	C	68	60.0	883	11	BC401991
69	C	69	60.0	899	11	BI161321
70	C	70	60.0	901	11	BC248223
71	C	71	60.0	915	13	CNS0421H
72	C	72	60.0	960	13	CNS02KRT
73	C	73	60.0	989	11	BE890850
74	C	74	60.0	1043	13	CNS0453G
75	C	75	60.0	1082	13	CNS03NN3
76	C	76	60.0	1111	11	BI246341
77	C	77	60.0	1163	11	BC752823
78	C	78	60.0	1860	11	BF342587
79	C	79	60.0	100	10	A1954642
80	C	80	60.0	119	11	BF483161
81	C	81	60.0	159	11	BF558381
82	C	82	60.0	160	10	AV249305
83	C	83	60.0	184	10	AM324941
84	C	84	60.0	202	10	BM418555
85	C	85	60.0	203	13	CNS0072V

BE144032	601791288
BE795069	601592885
W06020	T9EPT787f1
A1210255	h4a02a1.r
BF713183	MI-P-01-a
BC408765	gb76c07.y
AV537540	AV537540
BC382504	298426.MA
BF994483	CM3-GN010
BF402559	UI-R-CNO-
C72075	C72075.Rice
A1441334	sa55d11.y
C39749	C39749.Yoj1
AM699577	gb08f06.y
AL370675	MEB439C04
AL122961	Fugu.rubr
BE021456	sm49c06.y
A1938935	sc63b07.y
AQ205063	HS-3226_B
A1272538	uk05h06.y
AQ129442	HS-3045_A
AK017623	Mus.muscu
BC726684	sa31d02.
AO815087	HS-5261_B
A0066422	A0066422
A1507864	sa88b04.y
BF007983	1619492.A
AL379335	MEB44F11
BF634261	NF084G10D
AL303656	Tetraodon
AL381376	MEBC019F0
AO528128	RPCI-11-3
BI424083	sa05f03.
BC080788	h3056f11-
AO655910	Sheared.D
BC646892	EST08511
AM953979	EST165944
BC067650	H3056f11-
BC453113	NF092D6L
AM993297	RC2-BN003
BE821306	GM700024A
BE781869	601470202
AQ361397	mgx0004A
BE273024	601171345
BI161170	602865622
A1226887	uj10h02.y
AM959798	EST371869
BC122704	602351832
BC855614	1024043C0
BC081016	RPCI-24-3
ALA16695	T3.end.of
BF667013	963085F06
BI197672	602757580
AL254617	Tetraodon
AL275017	Tetraodon
BE890850	601431303
AL274885	Tetraodon
AL252264	Tetraodon
BI246341	602958985
BC752823	602732671
BF342587	602013717
A1954642	wq34e02.x
BF483161	WHE213_E
BF558381	UI-R-A1-e
AV249305	AV249305
AM324941	TENU4245
BM418555	BM418555
AL055653	Arabidops

```

c 86 13 52.0 208 10 A1329493
87 13 52.0 222 11 BF827222
88 13 52.0 225 10 A1973672
89 13 52.0 229 10 A1973672
90 13 52.0 235 11 BF84939
91 13 52.0 236 11 BF349456
92 13 52.0 238 10 AV321866
93 13 52.0 241 10 BF036694
94 13 52.0 244 10 BF049989
95 13 52.0 245 10 BF087031
96 13 52.0 246 10 A0444613
97 13 52.0 246 10 BF285294
98 13 52.0 250 10 BF010678
99 13 52.0 251 10 BF090429
100 13 52.0 253 10 AA603352

ALIGNMENTS

RESULT 1
LOCUS A0113702 300 bp mRNA EST 19-OCT-2000
DEFINITION A0113702 unpublished oligo-capped cDNA library Caenorhabditis
elegans cDNA clone YK107 3', mRNA sequence.
ACCESSION A0113702
VERSION A0113702.1 GI:10927269
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
AUTHORS 1 (bases 1 to 300)
; Rhabditiidae; Peloderinae; Caenorhabditis.
TITLE A complementary view of the C. elegans genome
JOURNAL Unpublished (2000)
COMMENT Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES
source
1..300
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone_lib="unpublished oligo-capped cDNA library"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="varied"

BASE COUNT 97 a 50 c 60 g 93 t

Query Match 64.0%; Score 16; DB 10; Length 300;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ttgcgctcgcgcag 19
|||||
Db 211 TTTCGATCCTCGTCAG 226

RESULT 2
LOCUS A0116641 300 bp mRNA EST 19-OCT-2000
DEFINITION A0116641 unpublished oligo-capped cDNA library Caenorhabditis
elegans cDNA clone YK747b3 3', mRNA sequence.
ACCESSION A0116641

```

```

VERSION A0116641.1 GI:10930208
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
AUTHORS 1 (bases 1 to 300)
; Rhabditiidae; Peloderinae; Caenorhabditis.
TITLE A complementary view of the C. elegans genome
JOURNAL Unpublished (2000)
COMMENT Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES
source
1..300
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone_lib="unpublished oligo-capped cDNA library"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="varied"

BASE COUNT 97 a 51 c 58 g 92 t 2 others

Query Match 64.0%; Score 16; DB 10; Length 300;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ttgcgctcgcgcag 19
|||||
Db 221 TTTCGATCCTCGTCAG 236

RESULT 3
LOCUS BG269235 695 bp mRNA EST 20-FEB-2001
DEFINITION L0-326773 Ice plant Lambda Uni-Zap XR expression library, 0 hours
NaCl treatment Mesembryanthemum crystallinum cDNA clone L0-3267 5',
mRNA sequence.
ACCESSION BG269235
VERSION BG269235.1 GI:12975075
KEYWORDS EST.
SOURCE common ice plant.
ORGANISM Mesembryanthemum crystallinum
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Caryophyllidae; Caryophyllales; Alzooceae; Mesembryanthemum.
1 (bases 1 to 695)
Cushman, J.C.
AUTHORS An expressed sequence tag database for the common ice plant,
TITLE Mesembryanthemum crystallinum
JOURNAL Unpublished (1997)
COMMENT Contact: Cushman JC
Department of Biochemistry
University of Nevada
MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcushman@unr.edu
PCR Primers
FORWARD: T7
BACKWARD: T3
Plate: L0-33 row: F column: 7
Seq primer: T3
High quality sequence stop: 350

```


FEATURES POLYA=NO.
Location/Qualifiers
1. .695
/organism="Mesembryanthemum crystallinum"
/db_xref="taxon:3544"
/clone="L0-3267"
/clone_lib="rice plant lambda Uni-Zap XR expression library
/0 hours NaCl treatment"
/tissue_type="leaf"
/dev_stage="Six week old"
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI"
BASE COUNT 221 a 148 c 158 g 168 t
ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 695;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tgcattctgcagtg 21
|||||
Db 104 TCGATCCTCGTCAGGT 119

RESULT 4
B1042497 331 bp mRNA EST 14-JUN-2001
LOCUS
DEFINITION RC5-OT0098-220101-011-F11 OT0098 Homo sapiens cDNA, mRNA sequence.
ACCESSION B1042497
VERSION B1042497.1 GI:14449123
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 331)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=RC5&cl2=RC5-OT0098-
220101-011-F11&cl3=2001-01-22&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 59
High quality sequence stop: 331.
Location/Qualifiers
1. .331
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="OT0098"
/dev_stage="Adult"
/note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue

BASE COUNT 54 a 100 c 111 g 66 t
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 331;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atctctgcagtg 23
|||||
Db 70 ATCCTCGTCAGTGC 84

RESULT 5
AW925187 431 bp mRNA EST 19-JUL-2000
LOCUS
DEFINITION W51.76.E02.b1.A002 Water-stressed 1 (W51) Sorghum bicolor cDNA,
mRNA sequence.
ACCESSION AW925187
VERSION AW925187.1 GI:8091013
KEYWORDS EST.
SOURCE sorghum.
ORGANISM sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 431)
Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and Pratt,
L.H.
An EST database from Sorghum: water-stressed plants
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 339
POLYA=NO.
Location/Qualifiers
1. .431
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Water-stressed 1 (W51)"
/note="Organ: Mix of 5-week old plants on days 7 & 8 after
water was withheld; Vector: Lambda Zap; Site_1: XhoI;
Site_2: EcoRI; The library was made from poly-A RNA in the
cloning vector lambda Zap II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT 98 a 135 c 106 g 92 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 431;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 gatctctgcagtg 22
|||||
Db 419 GATCCTCGTCAGTGC 405

RESULT 6
B1042853 454 bp mRNA EST 14-JUN-2001
LOCUS
DEFINITION RC5-OT0098-020201-012-F07 OT0098 Homo sapiens cDNA, mRNA sequence.

ACCESSION B1042853
 VERSION B1042853.1 GI:14449479
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 454)

REFERENCE
 AUTHORS Dias Neto, F., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663

TITLE
 JOURNAL Contact: Simpson A.J.G.
 MEDLINE Laboratory of Cancer Genetics
 COMMENT Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/ICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=RC5&cl2=RC5-OT0098-020201-012-f07&f3=2001-02-02&f4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 128.
 Location/Qualifiers
 1..454
 /organism="Homo sapiens"
 /db_xref="taxon:3606"
 /clone_lib="OT0098"
 /dev_stage="Adult"
 /note="Organ: ovary; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from OESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 97 a 140 c 132 g 81 t 4 others
 ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 454;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 atccctcagtgac 23
 |||||||||||||
 DB 427 ATCTCTGTCAGTGTC 413

RESULT 7
 BE442332
 LOCUS BE442332 519 bp mRNA EST 25-JUL-2000
 DEFINITION 925017G02.xl C. reinhardtii CC-2290, normalized, lambda Zap II
 BE442332
 BE442332.1 GI:9441847
 EST.
 Chlamydomonas reinhardtii.
 Chlamydomonas reinhardtii.
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 Chlamydomonadaceae; Chlamydomonas.
 1 (bases 1 to 519)
 Grossman, A., Davies, J., Federspiel, N., Harris, E., Iefevre, P., McDermott, J.P., Silflow, C., Stern, D. and Surzycki, R.

TITLE
 JOURNAL Analyses of the Chlamydomonas reinhardtii Genome: A Model,
 COMMENT Unicellular System for Analyzing Gene Function and Regulation in
 Vascular Plants; Project phase 2
 Unpublished (2000)
 Contact: Elizabeth H. Harris
 DCMB Box 91000
 Duke University
 Durham, NC 27708-1000, USA
 Tel: 919 613 8164
 Fax: 919 613 8177
 Email: chlamydeku.edu.

FEATURES
 source Location/Qualifiers
 1..519
 /organism="Chlamydomonas reinhardtii"
 /strain="CC-2290 wild type mt- S1 D2"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-2290, normalized, lambda Zap II"
 /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; This library was constructed by John Davies and Jeffrey McDermott. RNA was isolated from strain CC-2290 (Minnesota isolate of C. reinhardtii) grown to mid-log phase in TAP (acetate containing) medium in the light. PolyA mRNA was purified, and cDNA was synthesized and directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK-plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."
 BASE COUNT 137 a 137 c 121 g 124 t
 ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 519;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tccctcagtgac 24
 |||||||||||||
 DB 72 TCCCTGTCAGTGCA 86

RESULT 8
 A2726169
 LOCUS A2726169 542 bp DNA GSS 24-JAN-2001
 DEFINITION RPI-24-76J3.TV RPI-24 Mus musculus genomic clone RPI-24-76J3,
 A2726169
 A2726169.1 GI:12473564
 GSS.
 house mouse.
 SOURCE
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 542)
 Zhao, S., Nieman, W., Malek, J., Shatsman, S., Akintu, B., Levins, M., Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E., Russell, D., de Jong, P. and Fraser, C.M.
 Mouse BAC End Sequences from Library RPI-24
 Unpublished (1999)
 Other GSSs: RPI-24-76J3.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end

page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 76 row: J column: 3
 Seq primer: T7
 Class: BAC ends.

FEATURES
 source Location/Qualifiers
 1..542

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPC1-24-76J3"
 /clone_lib="RPC1-24"
 /sex="Male"

/cell_type="Spleen/Brain"
 /note="Vector: pTARBACK1. Site 1: BamHI. Site 2: BamHI.
 RPC1-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pTARBACK1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

BASE COUNT 152 a 126 c 132 g 132 t

ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 542;

Best local Similarity 100.0%; Pred. No. 69;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tcgctccgcgcagc 20

Db 149 TCGATCCTCGTCAGC 163

Db

RESULT 9
 AZ168021/c 543 bp DNA GSS 29-AUG-2000

LOCUS SP_0103.B1.F05_17A Strongylocentrotus purpuratus, purple sea urchin

DEFINITION clone Plate=103 Col=9 Row=L, DNA sequence.

ACCESSION AZ168021

VERSION AZ168021.1

KEYWORDS GSS.

SOURCE Strongylocentrotus purpuratus.

ORGANISM Strongylocentrotus purpuratus

Strongylocentrotus purpuratus

Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;

Echinoidea; Echinoidea; Echinacea; Echinoidea;

Strongylocentrotidae; Strongylocentrotus.

1 (bases 1 to 543)

Camaron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,

Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray

,G.A., Etlensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and

Hood,L.

A sea urchin genome project: Sequence scan, virtual map, and

additional resources

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)

COMMENT 20402566

Contact: Camaron, RA, Davidson, EH, Hood, L

Division of Biology 156-29

California Institute of Technology

Pasadena California 91125, USA

Tel: (626) 395-8421

Fax: (626) 793-3047

Email: acameron@caltech.edu

Plate: 103 row: L column: 9

Seq primer: T7

Class: BAC ends

High quality sequence stop: 543.

Location/Qualifiers

1..543

/organism="Strongylocentrotus purpuratus"

/db_xref="taxon:7668"

/clone="Plate=103 Col=9 Row=L"

/clone_lib="Strongylocentrotus purpuratus, purple sea

urchin, sperm genomic BAC library"

/note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli

BASE COUNT 149 a 116 c 116 g 157 t 5 others

ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 543;

Best local Similarity 100.0%; Pred. No. 69;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atccctgcagtgac 23

Db 115 ATCCTCGTCAGTGC 101

RESULT 10

LOCUS C98409

DEFINITION C98409 Rice panicle at flowering stage Oryza sativa cDNA clone

E0106.62, mRNA sequence.

ACCESSION C98409

VERSION C98409.1

KEYWORDS EST.

SOURCE Oryza sativa.

ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 658)

Sasaki,T. and Yamamoto,K.

Rice cDNA from panicle at flowering stage

Unpublished (1996)

Contact: Takuji Sasaki

National Institute of Agrobiological Resources

Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki

305-8602, Japan

Tel: 81-298-38-7441

Fax: 81-298-38-7468

Email: tsasaki@abrc.affrc.go.jp, URL: <http://rgp.dna.affrc.go.jp/>

PROJECT "RGP"

FEATURES

source Location/Qualifiers

1..658

/organism="Oryza sativa"

/strain="Nipponbare"

/db_xref="taxon:4530"

/clone="E0106.62"

/clone_lib="Rice panicle at flowering stage"

/dev_stage="flowering stage"

/note="Organ: panicle; Rice cDNA from panicle at flowering

stage"

BASE COUNT 161 a 159 c 155 g 179 t 4 others

ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 658;

Best local Similarity 100.0%; Pred. No. 70;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tccctgcagtgca 24

Db 177 TCCTCGTCAGTGC 191

RESULT 11

LOCUS AUI65518

DEFINITION AUI65518 Rice panicle at flowering stage Oryza sativa cDNA clone

E2045, mRNA sequence.

ACCESSION AUI65518

VERSION AUI65518.1

KEYWORDS EST.

SOURCE Oryza sativa.

ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE
 Spermatoxylata; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 668)
 AUTHORS
 Sasaki, T. and Yamamoto, K.
 TITLE
 Rice cDNA from panicle at flowering stage (2000)
 JOURNAL
 Unpublished (2000)
 COMMENT
 Contact: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
 305-8602, Japan
 Tel: 81-298-38-7441
 Fax: 81-298-38-7468
 Email: tsasaki@agr.affrc.go.jp, URL: http://rgrp.dna.affrc.go.jp/
 PROJECT = "RGP".

FEATURES

source
 Location/Qualifiers
 1..668
 /organism="Oryza sativa"
 /strain="Nipponbare"
 /db_xref="taxon:4530"
 /clone="E2045"
 /clone_lib="Rice panicle at flowering stage"
 /dev_stage="flowering stage"
 /note="Organ: panicle; Rice cDNA from panicle at flowering stage"
 BASE COUNT
 160 a 162 c 162 g 179 t 5 others
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Score 15; DB 10; Length 668;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tccctgcagctgca 24
 ||||||||||||
 Db 198 TCCTCGTCAGTGCA 212

RESULT 12
 A0094990 713 bp mRNA EST 30-JUN-2000
 LOCUS
 A0094990 Rice panicle (between 3cm to 10cm) Oryza sativa cDNA clone
 E41284, mRNA sequence.
 ACCESSION
 A0094990
 VERSION
 A0094990.1 GI:8857672
 KEYWORDS
 EST.
 SOURCE
 Oryza sativa.
 ORGANISM
 Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatoxylata; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 713)
 Sasaki, T. and Yamamoto, K.
 TITLE
 Rice cDNA from panicle (between 3cm to 10cm) (2000)
 JOURNAL
 Unpublished (2000)
 COMMENT
 Contact: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
 305-8602, Japan
 Tel: 81-298-38-7441
 Fax: 81-298-38-7468
 Email: tsasaki@agr.affrc.go.jp, URL: http://rgrp.dna.affrc.go.jp/
 PROJECT = "RGP".
 E41284_87.

FEATURES

source
 Location/Qualifiers
 1..713
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /clone="E41284"
 /clone_lib="Rice panicle (between 3cm to 10cm)"
 /tissue_type="panicle"
 /dev_stage="panicle (between 3cm to 10cm)"

BASE COUNT
 167 a 188 c 180 g 175 t 3 others
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Score 15; DB 10; Length 713;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tccctgcagctgca 24
 ||||||||||||
 Db 288 TCCTCGTCAGTGCA 302

RESULT 13
 BF144032/c 1096 bp mRNA EST 24-OCT-2000
 LOCUS
 60179128BF1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4022252 5',
 DEFINITION
 mRNA sequence.
 ACCESSION
 BF144032
 VERSION
 BF144032.1 GI:10983072
 KEYWORDS
 EST.
 SOURCE
 house mouse.
 ORGANISM
 Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1096)
 NIH-MGC http://mgc.nci.nih.gov/.
 TITLE
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL
 Unpublished (1999)
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LHM9278 row: c column: 21
 High quality sequence stop: 630.
 Location/Qualifiers
 1..1096
 /organism="Mus musculus"
 /strain="C57BL/6 J"
 /db_xref="taxon:10090"
 /clone="IMAGE:4022252"
 /clone_lib="NCI CGAP Lu30"
 /tissue_type="tumor, metastatic to mammary"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pCMV-SPORT6; Site: 1; NOTI;
 Site-2: Salt; transgenic model WNT-1, expression driven by
 MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
 dt. Library constructed by Life Technologies.
 Investigator providing samples: Gilbert Smith, NIH"

FEATURES

source
 Location/Qualifiers
 1..1096
 /organism="Mus musculus"
 /strain="C57BL/6 J"
 /db_xref="taxon:10090"
 /clone="IMAGE:4022252"
 /clone_lib="NCI CGAP Lu30"
 /tissue_type="tumor, metastatic to mammary"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pCMV-SPORT6; Site: 1; NOTI;
 Site-2: Salt; transgenic model WNT-1, expression driven by
 MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
 dt. Library constructed by Life Technologies.
 Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT
 336 a 271 c 271 g 218 t

Query Match
 Best Local Similarity 100.0%; Score 15; DB 11; Length 1096;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcaattgcagctcg 15
 ||||||||||||
 Db 1076 GCATTGCAGTCCTCG 1062

RESULT 14
 BE795069 1126 bp mRNA EST 20-SEP-2000
 LOCUS
 60159288BF1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3946869 5',
 DEFINITION
 mRNA sequence.
 ACCESSION
 BE795069

```

VERSION      BE795069.1  GI:10216267
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE    1 (bases 1 to 1126)
              NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS      National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE        Unpublished (1999)
JOURNAL      Contact: Robert Strausberg, Ph.D.
              Email: rs3apbs@femail.nih.gov
COMMENT      Tissue Procurement: DCM/D/DPF
              cDNA Library Preparation: Ling Hong/Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
              Plate: L1CM807 row: j column: 22
              High quality sequence start: 10
              High quality sequence stop: 702.
              Location/Qualifiers
                1..1126
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:3946869"
                /clone_lib="NIH-MGC_7"
                /tissue_type="small cell carcinoma"
                /cell_line="MGC3"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: lung; Vector: pOT87; Site_1: XhoI; Site_2:
                EcoRI; cDNA made by oligo-dT priming. Directionally
                cloned into EcoRI/XhoI sites using the following 5'
                adaptor: GGCACGAG(G). Size-selected >500bp for average
                insert size 1.8kb. Library constructed by Ling Hong in
                the laboratory of Gerald M. Rubin (University of
                California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT   262 a      320 c      289 g      255 t
ORIGIN
```

```

Query Match      60.0%; Score 15; DB 11; Length 1126;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

Qy      8 gattcctgcaggtg 22
        |||
Db      853 GATCCTGCTGACGCTG 839
```

```

RESULT 15
LOCUS      W06020      195 bp      mRNA      EST      22-MAY-2000
DEFINITION TgSTZy87f10.r1 TgRH Tachyzoite cDNA Toxoplasma gondii cDNA clone
            tgz87f10.r1 5', mRNA sequence.
ACCESSION  W06020
VERSION     W06020.1  GI:1278733
KEYWORDS   EST.
SOURCE     Toxoplasma gondii.
            Toxoplasma gondii.
ORGANISM   Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
            Sarcocystidae; Toxoplasma.
REFERENCE  1 (bases 1 to 195)
            Hehl, A., Manger, T., Marra, M., Sibley, L.D., Ajikwa, J.A., Aslett, M.A.,
            Dietrich, N., Dubnue, T., Hillier, L., Kucaba, T., Wan, K.L.,
            Waterston, R.H. and Boothroyd, J.
            Washoe-Merck-Stanford-NIH Toxoplasma EST project
            Unpublished (1996)
            Contact: Marra M
            Washoe-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
```

```

Tel: 314 286 1800
Fax: 314 286 1810
Email: toxo@watson.wustl.edu
David Sibley at toxoest@porcim.wustl.edu for further information
relating to organism, clone or library availability.
Seq primer: T3
High quality sequence stop: 148.
Location/Qualifiers
```

```

FEATURES
  source
```

```

1..195
/organism="Toxoplasma gondii"
/strain="RH"
/db_xref="taxon:5811"
/clone="tgzy87f10.r1"
/clone_lib="TgRH tachyzoite cDNA"
/lab_host="XLI-Blue MRF"
/note="vector: Lambda ZAP; Site_1: EcoRI; Site_2: XhoI;
Toxoplasma RH strain tachyzoites were grown in human
foreskin fibroblast cultures in vitro. The library was
constructed by K.L. Wan, Cambridge University. cDNAs were
synthesized from polyA RNAs by oligo d(T) priming and
directionally cloned into the EcoRI to XhoI sites of the
Lambda ZAPII vector using the ZAP-cDNA synthesis kit
(Stratagene). WARNING: the library contains a small
percentage of cDNAs derived from the human host cells."
```

```

BASE COUNT   59 a      45 c      62 g      27 t
ORIGIN
```

```

Query Match      56.0%; Score 14; DB 11; Length 195;
Best Local Similarity 100.0%; Pred. No. 2,5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

Qy      1 gatttcgattc 14
        |||
Db      181 GCATTTCGATCTC 168
```

```

RESULT 16
LOCUS      A1210255      196 bp      mRNA      EST      19-OCT-1998
DEFINITION h4a02a1.f1 Aspergillus nidulans 24hr asexual developmental and
            vegetative cDNA Lambda zap library Emericella nidulans cDNA clone
            h4a02a1 5', mRNA sequence.
ACCESSION  A1210255
VERSION     A1210255.1  GI:3772197
KEYWORDS   EST.
SOURCE     Emericella nidulans.
            Emericella nidulans
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; Emericella.
REFERENCE  1 (bases 1 to 196)
            Kupfer, D., Gray, J., Hausner, J., Lai, H., Martin, W., Aramayo, R.,
            Prade, R. and Roe, B.
            An Aspergillus nidulans EST Database
            Unpublished (1998)
            Other ESTs: h4a02a1.f1
            Contact: Bruce A. Roe, University of Oklahoma, broeou@ou.edu
            Department of Chemistry and Biochemistry
            Advanced Center for Genome Technology, University of Oklahoma
            620 Parrington Oval, Norman, OK 73019, USA
            Tel: 405 325 4912
            Fax: 405 325 7762
            Email: broeou@ou.edu
            We anticipate the future release of the cDNA clones to the Fungal
            Genetics Stock Center
            Seq primer: T3
            High quality sequence stop: 189.
            Location/Qualifiers
              1..196
              /organism="Emericella nidulans"
              /strain="FGSC A26"
              /db_xref="taxon:162425"
              /clone="h4a02a1"
```

/clone.lib="Aspergillus nidulans 24hr asexual developmental and vegetative cDNA lambda zap library" /tissue_type="vegetative mycelia, asexual structures" /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI; 5' end of cDNA cloned into EcoRI site of Bluescript; 3' end of cDNA cloned into XhoI site of Bluescript"

BASE COUNT 43 a 60 c 37 g 55 t 1 others

ORIGIN

Query Match 56.0%; Score 14; DB 10; Length 196;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 gatccctcagct 21
|||||

Db 32 GATCCTCCTCAGCT 45

RESULT 17
BF713183/c
LOCUS BF713183 208 bp mRNA EST 02-JAN-2001
DEFINITION MI-P-O1-adt-d-09-1-UM.s1 MI-P-O1 Sus scrofa cDNA clone
ACCESSION BF713183
VERSION BF713183.1 GI:12012658
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 208)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
JOURNAL MEDLINE
COMMENT Contact: Tugale CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildeer Hall, Ames, IA 50011-3150, USA
Tel: 515294252
Fax: 5152942401
Email: ctugale@iastate.edu
The sequence contained an oligo-dt track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dt track served to verify it as a clone from the non-normalized ovary at estrus day 0 library cDNA library
Preparation: RJ Woods, JA Green, RS Praher, SJ42 Animal Science Research Center, Department of Animal Science, University of Missouri-Columbia, 65211 Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 145-208, >g120877|emb1X64127.1|SSPR S.scrofa DNA for SINE sequence SSPRE
Seq primer: M13 Forward
POLYA=yes.

FEATURES
source
1..208
Location/Qualifiers
/organism="Sus scrofa"
/strain="crossbred"
/db_xref="taxon:9823"
/clone="MI-P-O1-adt-d-09-1-UM"
/clone.lib="MI-P-O1"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-O1 library is derived from ovary at estrus day 0. For a detailed description of the library from which this clone was derived, please visit our web site at
http://bigest.genome.iastate.edu/
TAG_LIB=MI-P-O1

TAG_TISSUE=ovary at estrus day 0
TAG_SEQ=GACGTA"

BASE COUNT 68 a 31 c 43 g 66 t

ORIGIN

Query Match 56.0%; Score 14; DB 11; Length 208;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 gatccctcagct 21
|||||

Db 189 GATCCTCCTCAGCT 176

RESULT 18
BG408765/c
LOCUS BG408765 257 bp mRNA EST 13-MAR-2001
DEFINITION gb76c07.y1 Moss EST library PPG Physcomitrella patens cDNA clone
ACCESSION BG408765
VERSION BG408765.1 GI:13315054
KEYWORDS EST.
SOURCE Physcomitrella patens.
ORGANISM Physcomitrella patens.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE 1 (bases 1 to 257)
Quatrano, R., Bashardes, S., Cove, D., Cumming, A., Knight, C., Clifton, S., Marra, M., Hillier, L., Pape, D., Martin, J., Wylie, T., Underwood, K., Theising, B., Allen, M., Bowers, Y., Person, B., Swallier, T., Steptoe, M., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R., Waterston, R. and Wilson, R.
Leeds/Wash U Moss EST Project
Unpublished (1999)
Contact: Ralph Quatrano
Leeds/Wash U Moss EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Libraries were constructed by Dr. Stavros Bashardes as part of the Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and Washington Univ. in St. Louis (USA) DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
Seq primer: -40RP from Glibco
High quality sequence stop: 239.

FEATURES
source
1..257
Location/Qualifiers
/organism="Physcomitrella patens"
/db_xref="taxon:3218"
/clone="PEP_SOURCCE_ID:PPG.CopyA-90613"
/clone.lib="Moss EST library PPG"
/tissue_type="gametophore: 30 day old tissue, ammonium-grown"
/lab_host="DH10B"
/note="Vector: PAMPI; Construction of the cDNA library was performed by Dr. W. Gregg Clark using a modification of the cDNA synthesis protocol developed in the laboratory of Dr. Michael Lovett by Dr. Yulia Korshunova (personal communication). First polyA + RNA was isolated from total gametophore RNA using oligo dt magnetic beads. Following this, first strand cDNA synthesis was performed on the bead-bound polyA + RNA, during which an oligonucleotide anchor sequence was incorporated onto the 5'-ends of the cDNA. PCR amplification was then used to synthesize the second strand, to amplify the double stranded DNA, and to incorporate drrp containing sequences into the ends of the double stranded cDNA. This DNA was size selected and cloned into PAMPI using the CloneAMP PAMPI System (Life Technologies, GibcoBRL) for cloning amplification products

by a non-restriction site dependant process. The cloning was directional based on sequence asymmetry introduced at the ends during PCR amplification. The 3' cDNA ends are proximal to the NotI site of the multiple cloning site in pAMP1. This annealing mixture was transformed into chemically competent DH10B cells and selected for ampicillin resistant growth. The resulting clones (about 330,000) were pooled to make the library."

BASE COUNT 52 a 50 c 76 g 79 t
ORIGIN

Query Match 56.0%; Score 14; DB 11; Length 257;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 gatccctgcagct 21
|||||
Db 201 GATCCTCGTCAGGT 188

RESULT 19
AV537540 266 bp mRNA EST 06-SEP-2000
LOCUS AV537540 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
DEFINITION CDNA clone R208c08f 3', mRNA sequence.

AV537540
AV537540.1 GI:8697823

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

EST.
thale cress.
Arabidopsis thaliana
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots;
Rosidae: eurosids II: Brassicales: Brassicaceae: Arabidopsis.

REFERENCE
AUTHORS
TITLE

1 (bases 1 to 266)
Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)

JOURNAL
MEDLINE
COMMENT

20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers

FEATURES
source
1. 266

/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="R208c08f"
/clone_1lb="Arabidopsis thaliana roots Columbia"
/tissue_type="roots"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 83 a 68 c 66 g 49 t
ORIGIN

Query Match 56.0%; Score 14; DB 10; Length 266;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 ttgcagctcagct 17
|||||
Db 157 TTTCGATCCTCGTC 144

RESULT 20
BG382504 288 bp mRNA EST 12-MAR-2001
LOCUS BG382504 288 bp mRNA EST 12-MAR-2001
DEFINITION 298426 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BG382504
BG382504.1 GI:13306976
EST.
pig.
Sus scrofa
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Cetartiodactyla: Suidae: Sus.
1 (bases 1 to 288)
REFERENCE
AUTHORS

Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.
and Keele, J.W.

Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine

JOURNAL
COMMENT

Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt. trimmed with phred
v0.980904.e. Vector identified by cross_match with the -mnscore 18
and mismatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCAGCAGC
Plate: 4 row: G column: 9
Seq primer: ATTTCAGCTGACACTATAG.

FEATURES
source
1. 288

/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_1lb="MARC 1PIG"
/tissue_type="pooled"
/db_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."

BASE COUNT 126 a 57 c 56 g 49 t
ORIGIN

Query Match 56.0%; Score 14; DB 11; Length 288;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 gatccctgcagct 21
|||||
Db 87 GATCCTCGTCAGGT 100

RESULT 21
BF994483 291 bp mRNA EST 23-JAN-2001
LOCUS BF994483 291 bp mRNA EST 23-JAN-2001
DEFINITION CM3-GN0102-031100-459-h10 GN0102 Homo sapiens cDNA, mRNA sequence.

BF994483
BF994483.1 GI:12400806
EST.
KEYWORDS
SOURCE
ORGANISM

human.
Homo sapiens
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.

REFERENCE
AUTHORS

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?rl=CM3<2=CM3-GN0102-
031100-459-h10<3=2000-11-03<4=1)
Seq primer: puc 18 forward
High quality sequence stop: 177.

FEATURES

source

Location/Qualifiers
1..291

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0102"
/dev_stage="Adult"

/note="Organ: placenta; normal; Vector: puc18; Site_1: Sma I
; Site_2: Sma I; A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the pUC 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."

BASE COUNT

66 a 80 c 68 g 77 t

ORIGIN

Query Match 56.0%; Score 14; DB 11; Length 291;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 gatccgtcaggt 21
|||||
DB 17 GATCCTGTCAGGT 30

RESULT 22

LOCUS

BF402559/c

DEFINITION

BF402559 304 bp mRNA EST 28-NOV-2000
UT-R-CA0-bhr-g-10-0-UI.s1 UT-R-CA0 Rattus norvegicus cDNA clone

ACCESSION

BF402559

VERSION

BF402559.1

KEYWORDS

EST.

SOURCE

Norway rat.

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 304)

Bonaldo, M.F., Lennon, G., and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene
discovery

Genome Res. 6 (3), 791-806 (1996)

97044477

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a cloneable poly A
tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

source

Location/Qualifiers
1..304

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"

/clone="01-R-CA0-bhr-g-10-0-UI"
/clone_lib="01-R-CA0"

/lab_host="DH10B (Life Technologies)"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The 01-R-CA0
library is a subtracted library derived from the following
tissues: thalamus, cerebellum, hypothalamus, medulla, pons
, midbrain, cerebral cortex, corpus striatum, testis, and
hippocampus. For a detailed description of the library
from which this clone was derived, please visit our web
site at ratest.eng.uiowa.edu. The subtraction has been
previously described in (Bonaldo, Lennon and Soares,
Genome Research 6:791-806, 1996)

BASE COUNT

75 a 68 c 73 g 88 t

ORIGIN

Query Match 56.0%; Score 14; DB 11; Length 304;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tctctgtcaggtgc 23
|||||
DB 246 TCTCTGTCAGGTGC 233

RESULT 23

LOCUS

C72075/c

DEFINITION

C72075 305 bp mRNA EST 22-SEP-1997
E0927_1A, mRNA sequence.

ACCESSION

C72075

VERSION

C72075.1

KEYWORDS

EST.

SOURCE

Oryza sativa.

ORGANISM

Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 305)

Sasaki, T. and Yamamoto, K.

Rice cDNA from panicle at flowering stage

Unpublished (1996)

Contact: Takuji Sasaki

National Institute of Agricultural Resources

Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki

305-8602, Japan

Tel: 81-298-38-7441

Fax: 81-298-38-7468

Email: tsasaki@agr.affrc.go.jp. URL: http://rpg.dna.affrc.go.jp/.

Location/Qualifiers

1..305

/organism="Oryza sativa"

/strain="Nipponbare"

/db_xref="taxon:4530"

/clone="E0927_1A"

/clone_lib="Rice panicle at flowering stage"

/dev_stage="flowering stage"

/note="Organ: panicle; Rice cDNA from panicle at flowering
stage"

BASE COUNT 30 a 127 c 80 g 57 t 11 others

ORIGIN

Query Match 56.0%; Score 14; DB 11; Length 305;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 cgaatccctgcagc 20
 |||
 Db 192 CATTCTGCTCAGC 179

RESULT 24
 A1441334 327 bp mRNA EST 18-APR-2000
 LOCUS sa55d11.y1 Gm-c1004 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 DEFINITION Gm-c1004-3214 5' similar to TR:Q43096 Q43096 CAFFEIC ACID
 O-METHYLTRANSFERASE ; mRNA sequence.

ACCESSION A1441334
 VERSION A1441334
 KEYWORDS A1441334.1 GI:4288448
 SOURCE soybean
 ORGANISM Glycine max

REFERENCE 1 (bases 1 to 327)
 AUTHORS Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna
 A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
 Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
 , R., Peterson, R., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
 , R., Rutter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
 , R., Watterston, R. and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 CONTACT: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu

ABSTRACT
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 for further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 Insert length: 1286 Std Error: 0.00
 Seq primer: 40bp from Gibco
 High quality sequence stop: 273
 POLY-A=NO.

FEATURES
 source
 1. 327
 Location/Qualifiers
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1004-3214"
 /clone_1lb="Gm-c1004"
 /tissue_type="root"
 /lab_host="XLI0-Cold"
 /note="Vector: pBluescript II XR; Site 1: EcoRI; Site 2:
 XhoI; Root cDNA. The mRNA was isolated from entire roots
 of 8 day old 'Williams' seedlings which were propagated on
 paper towels with distilled water. Stratagene's cDNA
 synthesis kit (catalog #200401) was used to synthesize the
 cDNA. First-strand synthesis was performed with 5-methyl
 dCTP, hence the ligated cDNA is hemimethylated.
 Stratagene's first-strand synthesis primer was used
 [GAGAGAGAGAGAGAGACTGCTCGAG(TT)18]. After
 second-strand synthesis, the cDNA ends were 'polished'
 with clone Pfu DNA polymerase, ligated to EcoRI adaptors,
 and phosphorylated. The XhoI site within the first-strand
 synthesis primer was restricted by digestion with XhoI;
 all XhoI sites in the cDNA would be protected by their
 hemimethylated status. The cDNA constructs were
 size-fractionated with a 500bp cutoff, using GibcoBRL Life
 Technologies' cDNA size fractionation column. The column
 eluent was then ligated into Stratagene's pBluescript II
 XR Predigested vector (pBluescript II SK(+)) that had been

BASE COUNT 97 a 73 c 63 g 94 t
 ORIGIN
 Query Match 56.0%; Score 14; DB 10; Length 327;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ctcgcagatgcag 25
 |||
 Db 188 CTCGTCAGTCTCAG 201

RESULT 25
 C39749 360 bp mRNA EST 18-OCT-1999
 LOCUS C39749 Yui Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
 DEFINITION Caenorhabditis elegans cDNA clone YK217c3 5' mRNA sequence.

ACCESSION C39749
 VERSION C39749
 KEYWORDS C39749.1 GI:2375986
 SOURCE EST.
 ORGANISM Caenorhabditis elegans.

REFERENCE 1 (bases 1 to 360)
 AUTHORS Kohara, Y., Motolashvili, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano
 , M., Miyata, A. and Nishigaki, A.
 Expression map of the C.elegans genome
 Unpublished (1996)
 CONTACT: Yui Kohara
 Genome Biology Lab.
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 81-559-81-6854
 Fax: 81-559-81-6855
 Email: ykohara@lab.nig.ac.jp.

FEATURES
 source
 1. 360
 Location/Qualifiers
 /organism="Caenorhabditis elegans"
 /strain="N2"
 /db_xref="taxon:6239"
 /clone="YK217c3"
 /clone_1lb="Yui Kohara unpublished cDNA:Strain N2
 hermaphrodite embryo"
 /sex="hermaphrodite"
 /dev-stage="embryo"

BASE COUNT 116 a 71 c 114 g 58 t 1 others
 ORIGIN
 Query Match 56.0%; Score 14; DB 11; Length 360;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cattcgatccgcg 15
 |||
 Db 105 CATTGATCTCTCG 92

RESULT 26
 AW69577/c 368 bp mRNA EST 18-APR-2000
 LOCUS AW69577/c

digested with EcoRI and XhoI, and phosphorylated). Both
 the white and blue colonies appear to contain recombinant
 plasmids with cDNA inserts. Blue colonies (9n=15) have been
 sequenced, and possess putative cDNA inserts. This library
 was constructed by Dr. Paul Keim & Virginia H. Coryell,
 Department of Biology, Box5640, Northern Arizona
 University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr.
 Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax:
 520-523-7500, email: paul.keim@nau.edu,
 virginia.coryell@nau.edu"

DEFINITION gb08f06.y1 Moss EST library PPN Physcomitrella patens cDNA clone
 PEPP_SOURCE_ID:PPN081112 5', mRNA sequence.
 ACCESSION AM699577
 VERSION AW699577.1 GI:7583670
 KEYWORDS EST.
 SOURCE Physcomitrella patens.
 ORGANISM Physcomitrella patens.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
 Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
 REFERENCE 1 (bases 1 to 368)
 Quatrano, R., Bashlades, S., Cove, D., Cuning, A., Knight, C., Clifton
 S., Maria, M., Hillier, L., Pape, D., Martin, J., Wylie, T., Underwood
 K., Theising, B., Allen, M., Bowers, V., Person, B., Swaller, T.,
 Steptoe, M., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,
 Waterston, R., and Wilson, R.
 Leeds/Wash U Moss EST Project
 TITLE Unpublished (1999)
 JOURNAL Contact: Ralph Quatrano
 COMMENT Leeds/Wash U Moss EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@washington.wustl.edu
 Libraries were constructed by Dr. Stavros Bashlades as part of the
 Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and
 Washington Univ. in St. Louis (USA) DNA sequencing by: Washington
 University Genome Sequencing Center for information on obtaining a
 clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
 Seq primer: 40RP from Gibco.
 FEATURES
 Source
 Location/Qualifiers
 1..368
 /organism="Physcomitrella patens"
 /db_xref="taxon:3218"
 /clone="PEP_SOURCE_ID:PPN081112"
 /clone_lib="Moss EST library PPN"
 /russue_type="protonemata: 7 day old tissue auxin treated"
 /lab_host="DH10B"
 /note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2:
 XhoI; Construction of the cDNA library was carried out
 using Stratagene's 'Unizap - cDNA synthesis kit'. cDNA was
 constructed using an oligo dT primer/linker that contains
 a XhoI site within it. Following ds cDNA synthesis,
 EcoRI adapters were ligated to the blunt ends and sample
 was digested with XhoI. The result is cDNA with an EcoRI
 sticky end on one side and a XhoI sticky end on the other.
 This cDNA was ligated directionally in Unizap arms. The
 vector is designed containing the pBluescript sequence as
 well as lambda DNA and cDNA is cloned within this
 pBluescript sequence. The vector was then packaged using
 Gold gigaPackaging extracts. Library was grown in XL1Blue
 MRF' cells and amplified. The library was excised by mass
 excision using Stratagene's 'Mass excision kit' that uses
 exsist as a helper phage that releases the pBluescript
 sequence and circularises it as single stranded plasmids
 that are then packaged (by helper phage) and secreted out
 of the host cell as phagemids. SOUR cells were transformed
 with phagemids and the library was plated out on LB-amp
 plates to select for transformants. Approximately 1,000
 ,000 colonies were grown and recovered. The double
 stranded plasmid library was recovered by using Qiagen
 Midi prep kit. 2 micro grams of each library were used to
 transform DH10B cells by electroporation."

BASE COUNT 81 a 65 c 98 g 124 t
 ORIGIN

Query Match 56.0%; Score 14; DB 10; Length 368;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 8 gatcctgcagct 21
 |||

Db 105 GATCCTGCTCAGCT 92
 RESULT 27
 LOCUS AL370675/c 378 bp mRNA EST 03-AUG-2000
 DEFINITION MCB39C04R1 MIRA Medicago truncatula cDNA clone MCB39C04 T7, mRNA
 sequence.
 ACCESSION AL370675
 VERSION AL370675.1 GI:9670428
 KEYWORDS EST.
 SOURCE Medicago truncatula
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
 Medicago.
 REFERENCE 1 (bases 1 to 378)
 Journet, E.P., Crespeau, H., van-Tuinen, D., Gouzy, J., Jallion, O.,
 Niebel, A., Carreau, V., Chatagnier, O., Kahn, D., Gianinazzi-Pearson
 V. and Gamas, P.
 Medicago truncatula ESTs from nitrogen-starved roots
 TITLE Unpublished (2000)
 JOURNAL Contact: Genoscope
 COMMENT Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
 Biologie Molculaire des Relations Plantes-Microorganismes,
 CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
 Mt-est@toulouse.inra.fr Website :
 http://sequence.toulouse.inra.fr/Mtruncatula.html).
 FEATURES
 Source
 Location/Qualifiers
 1..378
 /organism="Medicago truncatula"
 /cultivar="Jemalong"
 /db_xref="taxon:3880"
 /clone="MIRA39C04"
 /clone_lib="MIRA"
 /russue_type="root tips"
 /dev_stage="harvested after 3 days of N-starvation"
 /note="Vector: pBluescript PSK; Site.1: EcoRI; Site.2:
 XhoI; Plants were grown in an aeroponic chamber for 14
 days on nitrogen-rich medium followed by 3 days on N-free
 medium. RNA was extracted from root tips (1-3 cm). cDNA
 was prepared from polyA+ enriched RNA. The cDNA was
 directionally ligated into Uni-zapXR vector from
 Stratagene and packaged using GigaPack Gold packaging
 extracts. Plasmids containing cDNA inserts were
 mass-excised from phage stocks using Exsist helper phage
 and propagated in SOUR cells. Clone ordering and
 sequencing was performed by the Centre National de
 Sequencage (Genoscope, Evry, France)."

BASE COUNT 132 a 60 c 75 g 111 t
 ORIGIN

Query Match 56.0%; Score 14; DB 10; Length 378;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 9 acccgcagctg 22
 |||
 Db 41 ATCTCTGCTCAGCTG 28
 RESULT 28
 LOCUS FR0035439 381 bp DNA GSS 22-OCT-1999
 DEFINITION Fugu rubripes GSS sequence, clone 018F14AD12, genomic survey
 sequence.
 ACCESSION AL122961
 VERSION AL122961.1 GI:6104576

KEYWORDS GSS: genome survey sequence.

SOURCE Takifugu rubripes.

ORGANISM Takifugu rubripes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu.

REFERENCE 1 (bases 1 to 381)
Elgar, G., Clark, M.S., Smith, S., Meek, S., Warner, S., Edwards, Y.J.K., Umranta, Y., Williams, G. and Brenner, S.

TITLE Direct Submission

JOURNAL Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, CB10 1SB. UK Email: biobehringmp.mrc.ac.uk

COMMENT Vector: pBluescript II KS
V-type: phagemid
PRIMER: KS
DESCR: One pass dye-terminator sequencing of cosmid cloned genomic sequence.

FEATURES Location/Qualifiers
1..381
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone_lib="cosmid 018f14"
/clone="018f14D12"

BASE COUNT 81 a 83 c 87 g 127 t 3 others

ORIGIN

Query Match 56.0%; Score 14; DB 13; Length 381;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tgcattctcgtcag 19
|||||

Db 148 TCGATCTCTGTCAG 161

RESULT 29
BE021456 420 bp mRNA EST 21-NOV-2000
LOCUS sm49c06.y1 Gm-cl028 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-cl028-7235 5' similar to TR:022555 O-METHYLRANSFERASE. 12
1 TR:022308 ;, mRNA sequence.

ACCESSION BE021456
VERSION BE021456
KEYWORDS EST.

SOURCE soybean.
Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE 1 (bases 1 to 420)
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shih, W., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Willson, R.

TITLE Public Soybean EST Project

JOURNAL Unpublished (1999)

COMMENT Contact: Shoemaker R/Public soybean EST project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)

427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Insert length: 1255 Std Error: 0.00
High quality sequence stop: 286.

FEATURES Location/Qualifiers
1..420
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl028-7235"
/clone_lib="Gm-cl028"
/clone_type="roots of 'superpod' plants"
/lab_host="DH10B"
/note="Vector: pBluescript II XR; Site 1: EcoRI; Site 2: XhoI; The mRNA was isolated from roots of glycine max 'superpod' plants generously donated by Dr. Gary Stacey. The seedlings were inoculated with Bradyrhizobium japonicus, strain USDA110 prior to harvest. StrataGene's cDNA synthesis kit (catalog number 200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA was hemimethylated. A modification of StrataGene's first-strand synthesis primer was used. An 'anchor' nucleotide (V-A, C, or G) was added to the 3' end of the primer [GAGAGAGAGAGAGAGAGACTGCTCGAG(T)18V] to anchor the primer at the 5' end of the poly(A) tract. After second-strand synthesis, the cDNA ends were filled in with cloned pfu DNA polymerase, ligated to EcoRI adapters and subsequently phosphorylated. The XhoI site within the first-strand synthesis primer was then restricted by digestion with XhoI; all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500bp cutoff, using GibcoBRL Life Technologies' cDNA size fractionation column. The column eluent was then ligated into StrataGene's pBluescript II XR predigested vector (pBluescript II SK(+)) that has been digested with EcoRI and XhoI, and phosphorylated by StrataGene. Both the white and blue colonies appear to contain recombinant plasmids with cDNA inserts, based on size (n=25). This library was constructed by Dr. Paul Keim and Dr. Virginia Coryell."

BASE COUNT 124 a 92 c 82 g 122 t

ORIGIN

Query Match 56.0%; Score 14; DB 10; Length 420;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ctgcacagtcag 25
|||||

Db 221 CTCGTCAGTCGAC 234

RESULT 30
A1938935 421 bp mRNA EST 17-JUL-2000
LOCUS sc63b07.y1 Gm-cl016 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-cl016-1070 5' similar to SW:SPRI_ARATH 022315 PRE-MRNA SPLICING
FACTOR SP2 ;, mRNA sequence.

ACCESSION A1938935
VERSION A1938935
KEYWORDS EST.

SOURCE soybean.
Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE 1 (bases 1 to 421)
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers

TITLE
JOURNAL
COMMENT

, Y., Peterson, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 for further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Insert Length: 696 Std Error: 0.00
High quality sequence stop: 406.

FEATURES

source

Location/Qualifiers
1. 421
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl016-1070"
/clone_1lb="Gm-cl016"
/tissue.type="Immature flowers of field grown plants"
/lab.host="XLI0-gold"
/note="vector: pluescript II XR; Site1: EcoRI; Site2: XhoI; This cDNA library was constructed from mRNA isolated from immature flowers of field grown plants. The cDNA library was prepared using the Stratagene pluescript II XR library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pluescript vector. The ligated cDNA fragments were transformed into XLI0-gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelting."

BASE COUNT 111 a 81 c 114 g 115 t
ORIGIN

Query Match 56.0%; Score 14; DB 10; Length 421;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 tctctcagatgc 23
|||||
db 402 tctctcagatgc 415

RESULT 31
AO205063/c 429 bp DNA GSS 17-SEP-1998
LOCUS HS_3226_B2_E12_MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=3226 Col=24 Row=J, DNA sequence.
ACCESSION AO205063
VERSION AO205063.1 GI:3615633
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 429)
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, D., Young, D., Zhao, S., Adams, M.D. and
Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE
COMMENT

99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3226 row: J column: 24
Class: BAC ends
High quality sequence stop: 429.

FEATURES

source

Location/Qualifiers
1. 429
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3226 Col=24 Row=J"
/clone_1lb="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 120 a 84 c 71 g 153 t 1 others
ORIGIN

Query Match 56.0%; Score 14; DB 13; Length 429;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 cctctcagatgc 24
|||||
db 338 cctctcagatgc 325

RESULT 32
AI272538/c 430 bp mRNA EST 18-NOV-1998
LOCUS UK05406.y1 Schiller mouse MAC13 Mus musculus cDNA clone
DEFINITION IMAGE:1958555.5, similar to gb:537431 LAMININ RECEPTOR (HUMAN);
gb:02870 Mouse laminin receptor mRNA, complete cds (MOUSE);, mRNA
sequence.

ACCESSION AI272538
VERSION AI272538
KEYWORDS AI272538.1 GI:3894806
SOURCE EST.
ORGANISM house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 430)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMIT Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMIT Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMGE Consortium (info@imge.llnl.gov) for further information.
MGI:984895
Trace considered overall poor quality
Seq primer: Primer name ambiguous
High quality sequence stop: 1.

REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT

FEATURES
source
1. 430
/organism="Mus musculus"
/db_xref="taxon:10090"

digested with EcoRI and XhoI, and phosphorylated). Both the white and blue colonies appear to contain recombinant plasmids with cDNA inserts. Blue colonies (n=15) have been sequenced, and possess putative cDNA inserts. This library was constructed by Dr. Paul Keim & Virginia H. Coryell, Department of Biology, Box 5640, Northern Arizona University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr. Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax: 520-523-7500, email: paul.keim@nau.edu, virginia.coryell@nau.edu

BASE COUNT 144 a 103 c 101 g 145 t 1 others
ORIGIN

Query Match 56.0%; Score 14; DB 10; Length 494;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ctctgacgtacag 25
|||||
Db 211 CTGTCAGTGTCAG 224

RESULT 39
BF007983 518 bp mRNA EST 06-OCT-2000
LOCUS 1619492 Amblyomma americanum adult lambda Zap Express Amblyomma
DEFINITION americanum cDNA, mRNA sequence.
ACCESSION BF007983
VERSION BF007983.1 GI:10708258
KEYWORDS EST.
SOURCE Amblyomma americanum.
ORGANISM Amblyomma americanum.

REFERENCE
AUTHORS Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
Parasitiformes; Ixodida; Ixodidae; Amblyomma.
TITLE Hill, C.A. and Gutierrez, J.A.
1 (bases 1 to 518)
Analysis of the expressed genome of the lone star tick, *Amblyomma americanum* (Acari: Ixodidae) using an expressed sequence tag approach

JOURNAL
COMMENT Microb. Comp. Genomics 5 (2), 89-101 (2000) In press
Contact: Hill CA
Animal Science Discovery Research
Elanco Animal Health, A Division of Eli Lilly and Company
PO Box 708, 2001 West Main St., Greenfield, IN 46140, USA
Tel: 317 277 0826
Fax: 317 277 4522
Email: HILL.CATHERINE.ATLILLY.COM.

FEATURES
Source Location/Qualifiers
1..518
/organism="Amblyomma americanum"
/db_xref="taxon:6943"
/clone_lib="Amblyomma americanum adult lambda Zap Express"
/sex="Male, Female"
/dev_stage="Adult"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 69 a 164 c 148 g 134 t 3 others
ORIGIN

Query Match 56.0%; Score 14; DB 11; Length 518;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ctctgacgtacag 25
|||||
Db 27 CTGTCAGTGTCAG 40

RESULT 40
AL379335 521 bp mRNA EST 03-AUG-2000
LOCUS AL379335/c
DEFINITION MEBB44F11R1 MBBB Medicago truncatula cDNA clone MBB44F11 t77, mRNA

ACCESSION
VERSION AL379335.1 GI:9679087
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula

REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids f; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
1 (bases 1 to 521)
Journet, E.P., Crespeau, H., van-Tulnen, D., Gouzy, J., Jallion, O.,
Niebel, A., Carreau, V., Chatagnier, O., Kahn, D., Giaminazzi-Pearson
V. and Gamas, P.
Medicago truncatula ESTs from Sinorhizobium meliloti-induced root
nodules
Unpublished (2000)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
Biologie Molculaire des Relations Plantes-Microorganismes,
CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
M-est@toulouse.inra.fr Website :
http://sequence.toulouse.inra.fr/Mtruncatula.html).

FEATURES
Source Location/Qualifiers
1..521
/organism="Medicago truncatula"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone_lib="MtrBB44F11"
/clone_lib="MtrBB"
/tissue_type="symbiotic root nodules"
/dev_stage="harvested 4 days post inoculation with
Sinorhizobium meliloti"
/note="Vector: pBluescript PSK; Site_1: EcoRI; Site_2:
XhoI; Plants were grown in an aeroponic chamber on
nitrogen-rich medium for 21 days. Three days before
inoculation with Sinorhizobium meliloti, the medium was
replaced by N-free medium. Root nodules (+ short adjacent
root segments) were harvested 4 days post inoculation.
cDNA was prepared from polyA+ enriched RNA. The cDNA was
directionally ligated into Uni-Zap XR vector from
Stratagene and packaged using Gigapack Gold packaging
extracts. Plasmids containing cDNA inserts were
mass-excised from phage stocks using ExSact helper phage
and propagated in SOLR cells. Clone ordering and
sequencing was performed by the Centre National de
Sequencage (Genoscope, Evry, France)."

BASE COUNT 168 a 92 c 104 g 157 t
ORIGIN

Query Match 56.0%; Score 14; DB 10; Length 521;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atctctgacgtacg 22
|||||
Db 186 ATCTCTGTCAGTGTG 173

RESULT 41
BF634261 524 bp mRNA EST 19-DEC-2000
LOCUS BF634261/c
DEFINITION NF084G10DTF1083 Drought Medicago truncatula cDNA clone NF084G10DT
5', mRNA sequence.
ACCESSION BF634261
VERSION BF634261.1 GI:11898419
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula

Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eustosids I: Fabales: Fabaceae: Papilionoideae: Trifoliaceae: Medicago.

REFERENCE 1 (bases 1 to 524)
 Authors: Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores, H.R., Imman, J.T., Weller, J.W. and May, G.D.
 Title: Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula drought library
 Journal: Unpublished (2000)
 Comment: Contact: May GD
 The Samuel Roberts Noble Foundation
 2510 Sam Noble Parkway, Ardmore, OK 73402, USA
 Tel: 580 221 7391
 Fax: 580 221 7380
 Email: gdmay@noble.org
 Insert Length: 524 Std Error: 0.00
 Plate: 084 row: 5 column: 10
 Seq primer: TCACACAGCAACACGCTTGAC.

FEATURES
 source
 1..524
 /organism="Medicago truncatula"
 /db_xref="taxon:3880"
 /clone="MF084G10P"
 /clone_lib="Drought"
 /tissue_type="Plantlets"
 /dev_stage="Pooled timepoints"
 /note="Vector: Lambda Zap; Contains a mixture of entire plantlets harvested in a series of days-post-watering timepoints."

BASE COUNT
 133 a 126 c 102 g 162 t 1 others

ORIGIN

Query Match 56.0%; Score 14; DB 11; Length 524;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 524 ATCTCTCAGTGTG 511

RESULT 42
 CDS04RAN/C 532 bp DNA GSS 24-MAY-2000
 LOCUS Tetradon nigroviridis genome survey sequence T7 end of clone
 DEFINITION 003119 of library H from Tetradon nigroviridis, genomic survey sequence.
 ACCESSION AL303656
 VERSION AL303656.1 GI:8187784
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetradon nigroviridis.
 ORGANISM Tetradon nigroviridis
 Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi: Actinopterygii: Neopterygii: Teleostei: Euteleostei: Neoteleostei: Acanthomorphi: Acanthopterygii: Perciformes: Tetraodoniformes: Tetraodontidae: Tetradon.
 1 (bases 1 to 532)
 Roest-Crolius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.
 Title: Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis
 Journal: Unpublished
 2 (bases 1 to 532)
 Roest-Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F., Saurin, W. and Weissenbach, J.
 Title: Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence
 Journal: Unpublished

REFERENCE 3 (bases 1 to 532)
 Authors: Genoscope.
 Title: Direct Submission
 Journal: Submitted (12-APR-2000) to the EMBL/GenBank/DDJ databases
 Comment: This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetradon>.

FEATURES
 source
 1..532
 /organism="Tetradon nigroviridis"
 /db_xref="taxon:99883"
 /clone="003119"
 /clone_lib="H"
 /note="Genoscope sequence ID: C0BH003AE10XD1-end : T7"

BASE COUNT
 140 a 108 c 138 g 144 t 2 others

ORIGIN

Query Match 56.0%; Score 14; DB 13; Length 532;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 506 TTTCGATCTCTGTC 493

RESULT 43
 AL381376 548 bp mRNA EST 03-AUG-2000
 LOCUS MRC019F07R1 MRC Medicago truncatula cDNA clone MRC019F0 T7, mRNA
 DEFINITION sequence.
 ACCESSION AL381376
 VERSION AL381376.1 GI:9681127
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eustosids I: Fabales: Fabaceae: Papilionoideae: Trifoliaceae: Medicago.
 1 (bases 1 to 548)
 Journet, E.P., Crespeau, H., van Thunen, D., Gouzy, J., Jaillon, O., Niebel, A., Carreau, V., Chataigner, O., Kahn, D., Glanmazzi-Pearson, V. and Gamas, P.
 Title: Medicago truncatula ESTs from endomycorrhizal roots
 Journal: Unpublished (2000)
 Comment: Contact: Genoscope
 BP 191 91006 EVRY cedex - France
 Genoscope - Centre National de Sequencage
 Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
 Contact: Pascal Gamas and Etienne-Pascal Journet, Laboratoire de Biologie Moleculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email: Mc-est@toulouse.inra.fr Website: <http://sequence.toulouse.inra.fr/Mtruncatula.html>).

FEATURES
 source
 1..548
 /organism="Medicago truncatula"
 /cultivar="Jemalong"
 /db_xref="taxon:3880"
 /clone="MRC019F0"
 /clone_lib="MRC"
 /tissue_type="arbuscular mycorrhiza"
 /dev_stage="harvested 3 weeks post inoculation with Glomus intraradices"
 /note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2: XhoI; M. truncatula sterilised seeds were germinated for 72h at 25 C, before transplanting into a 1/3 Epsilones soil : 2/3 calined Terragreen mix in the presence of onion root fragments colonized by the arbuscular mycorrhizal fungus Glomus intraradices (Schenck & Smith, isolate LP8

Query Match
Best Local Similarity 100.0%; Score 14; DB 10; Length 548;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

168 a 101 c 112 g 167 t

fungus origin"

101 c 112 g 167 t

Query Match
Best Local Similarity 100.0%; Score 14; DB 10; Length 548;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

168 a 101 c 112 g 167 t

fungus origin"

101 c 112 g 167 t

Query Match
Best Local Similarity 100.0%; Score 14; DB 10; Length 548;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

168 a 101 c 112 g 167 t

fungus origin"

101 c 112 g 167 t

Query Match
Best Local Similarity 100.0%; Score 14; DB 10; Length 548;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

168 a 101 c 112 g 167 t

fungus origin"

101 c 112 g 167 t

Query Match
Best Local Similarity 100.0%; Score 14; DB 10; Length 548;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

168 a 101 c 112 g 167 t

fungus origin"

101 c 112 g 167 t

Query Match
Best Local Similarity 100.0%; Score 14; DB 10; Length 548;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

168 a 101 c 112 g 167 t

fungus origin"

101 c 112 g 167 t

Query Match
Best Local Similarity 100.0%; Score 14; DB 13; Length 559;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

168 a 101 c 112 g 167 t

fungus origin"

101 c 112 g 167 t

Query Match
Best Local Similarity 100.0%; Score 14; DB 13; Length 559;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

168 a 101 c 112 g 167 t

fungus origin"

101 c 112 g 167 t

Query Match
Best Local Similarity 100.0%; Score 14; DB 13; Length 559;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

168 a 101 c 112 g 167 t

fungus origin"

101 c 112 g 167 t

Query Match
Best Local Similarity 100.0%; Score 14; DB 13; Length 559;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

168 a 101 c 112 g 167 t

fungus origin"

101 c 112 g 167 t

Query Match
Best Local Similarity 100.0%; Score 14; DB 13; Length 559;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

168 a 101 c 112 g 167 t

fungus origin"

101 c 112 g 167 t

Query Match 56.0%; Score 14; DB 11; Length 579;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 10 tctctgctcaggtgc 23
|||||
Db 292 tctctgctcaggtgc 279

Search completed: December 26, 2001, 14:10:35
Job time: 7082 sec

1
2
3

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:53:07 ; Search time 1757.07 Seconds
(without alignments)
234.726 Million cell updates/sec

Title: US-09-396-196f-9

Perfect score: 25
Sequence: 1 cgatccctgcagtcagtcagtcagc 25

Scoring table: OLIGO_NUC
Gapex 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 9

Total number of hits satisfying chosen parameters: 220224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_hng: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_om: *
20: em_or: *
21: em_ov: *
22: em_pat: *
23: em_ph: *
24: em_pl: *
25: em_ro: *
26: em_sy: *
27: em_sts: *
28: em_un: *
29: em_vi: *
30: em_hng_hum: *
31: em_hng_in: *
32: em_hng_rod: *
33: em_hng_hum: *
34: em_hng_in: *
35: em_hng_rod: *
36: em_hng_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

8

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	1041	6	AR029499	AR029499 Sequence
2	25	100.0	1041	6	AR034916	AR034916 Sequence
3	25	100.0	1084	6	AL11530	AL11530 B108 gene o
4	25	100.0	1121	6	E00893	E00893 Genomic DNA
5	25	100.0	5793	1	ECOBIO	J04423 E.coli 7.8-
6	25	100.0	5872	6	A38246	A38246 Sequence 1
7	25	100.0	5872	6	A38251	A38251 Sequence 6
8	25	100.0	5872	6	A93674	A93674 Sequence 1
9	25	100.0	5872	6	A93679	A93679 Sequence 6
10	25	100.0	5872	6	AR101809	AR101809 Sequence
11	25	100.0	5872	6	AR101810	AR101810 Sequence
12	25	100.0	11022	1	AE005258	AE005258 Escherich
13	25	100.0	13501	1	AE005258	AE005258 Escherich
14	25	100.0	297816	1	AP002553	AP002553 Escherich
15	18	72.0	7215	1	SMABIO	D17468 S. marcesc
16	17	68.0	7318	10	AB026497	AB026497 Mus muscu
17	17	68.0	69900	2	AC021767	AC021767 Mus muscu
18	17	68.0	151340	2	AC069132	AC069132 Homo sapi
19	17	68.0	161742	2	AC068100	AC068100 Homo sapi
20	17	68.0	168360	2	AC041048	AC041048 Homo sapi
21	17	68.0	176379	2	AC060833	AC060833 Homo sapi
22	17	68.0	220166	2	AC022781	AC022781 Mus muscu
23	17	68.0	222212	2	AL591065	AL591065 Mus muscu
24	16	64.0	10536	1	AE004050	AE004050 Xylella f
25	16	64.0	71230	1	AC004508	AC004508 Homo sapi
26	16	64.0	155628	2	AC027810	AC027810 Homo sapi
27	16	64.0	175588	2	AL139326	AL139326 Human DNA
28	16	64.0	183700	9	AC005412	AC005412 Homo sapi
29	16	64.0	213732	1	AE001862	AE001862 Deinococc
30	16	64.0	294800	1	SMES91789	AL591789 Sinorhizo
31	16	64.0	313064	9	HSMA1B	AL442167 Homo sapi
32	16	64.0	340000	9	HS21C085	AL163285 Homo sapi
33	15	60.0	128	1	ECOBIOB	M27731 Escherichia
34	15	60.0	636	9	AF072252	AF072252 Homo sapi
35	15	60.0	728	8	TCU57884	U57884 Thanaeophor
36	15	60.0	1552	5	XLNFM4R	Z37526 X.laeviis HN
37	15	60.0	1727	9	AB056427	AB056427 Macaca fa
38	15	60.0	1801	8	RICCYTALD	D13512 Oryza sativ
39	15	60.0	2081	10	HAMHRP	M96650 Mesocricetu
40	15	60.0	2931	5	DRMJ5029	AJ005029 Danio rer
41	15	60.0	3829	9	HSMETHYL7	AF120994 Homo sapi
42	15	60.0	4380	9	HSAB17970	AJ317970 Homo sapi
43	15	60.0	4674	9	AF034373	AF034373 Homo sapi
44	15	60.0	4886	9	AB037816	AB037816 Homo sapi
45	15	60.0	5526	1	AF250776	AF250776 Unculture
46	15	60.0	6372	6	AX173175	AX173175 Sequence
47	15	60.0	8326	5	AF237379	AF237379 Xenopus l
48	15	60.0	8482	9	AF281036	AF281036 Homo sapi
49	15	60.0	12542	1	AE004708	AE004708 Pseudomon
50	15	60.0	20780	2	AC014847	AC014847 Drosophill
51	15	60.0	38065	1	MLCB1450	AL035159 Mycobacte
52	15	60.0	39520	2	AC091120	AC091120 Leishmani
53	15	60.0	44706	2	AC022328	AC022328 Mus muscu
54	15	60.0	50149	2	AC011955	AC011955 Homo sapi
55	15	60.0	59999	9	AL49627	AL49627 Human DNA
56	15	60.0	61010	9	AL138961	AL138961 Human DNA
57	15	60.0	73167	9	AC010316	AC010316 Homo sapi
58	15	60.0	85633	9	AC002088	AC002088 Human BAC
59	15	60.0	92388	9	HS1MDL	X97051 DNA sequenc
60	15	60.0	106482	10	AC007585	AC007585 Mus muscu
61	15	60.0	109973	2	AC007555	AC007555 Homo sapi
62	15	60.0	110000	2	AC087782_2	Continuation (3 of
63	15	60.0	119749	2	AC092845	AC092845 Homo sapi
64	15	60.0	122914	10	AC069451	AC069451 Mus muscu
65	15	60.0	137070	9	AC073052	AC073052 Homo sapi
66	15	60.0	136804	9	AP000469	AP000469 Homo sapi
67	15	60.0	141704	2	AC011754	AC011754 Homo sapi
68	15	60.0	144440	2	AC037436	AC037436 Homo sapi
69	15	60.0	150336	9	HS1395C13	AL117344 Human DNA
70	15	60.0	154804	2	AC022388	AC022388 Homo sapi

71 15 60.0 156119 2 AC073247 Homo sapi
72 15 60.0 157090 2 AB019441 Homo sapi
73 15 60.0 162043 2 AC040930 Homo sapi
74 15 60.0 164824 2 AC023887 Homo sapi
75 15 60.0 165026 2 AP000759 Homo sapi
76 15 60.0 165484 2 AL590671 Homo sapi
77 15 60.0 166353 2 AC091934 Homo sapi
78 15 60.0 168306 3 AC008135 Homo sapi
79 15 60.0 168972 2 AP003452 Homo sapi
80 15 60.0 169572 2 AC023832 Homo sapi
81 15 60.0 169802 2 CNS01D72 Homo sapi
82 15 60.0 175459 2 AC068538 Homo sapi
83 15 60.0 176947 9 HS170A21 Homo sapi
84 15 60.0 177085 3 AC008285 Homo sapi
85 15 60.0 177753 2 AP003123 Homo sapi
86 15 60.0 180568 2 AC021325 Homo sapi
87 15 60.0 182041 2 AC023025 Homo sapi
88 15 60.0 182396 2 AC092679 Homo sapi
89 15 60.0 182557 2 AC091942 Homo sapi
90 15 60.0 182599 2 AC022897 Homo sapi
91 15 60.0 184288 2 AC037427 Homo sapi
92 15 60.0 185255 9 AC090666 Homo sapi
93 15 60.0 185492 3 AC025339 Homo sapi
94 15 60.0 186893 2 AC026739 Homo sapi
95 15 60.0 187201 2 AC068063 Homo sapi
96 15 60.0 191006 2 AP001107 Homo sapi
97 15 60.0 191934 9 AC010729 Homo sapi
98 15 60.0 193530 2 AP003455 Homo sapi
99 15 60.0 195405 2 AC093355 Homo sapi
c 100 15 60.0 195621 2 AC021714 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
source location/Qualifiers
1..1041
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.00061;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctcgtagtgtagtcagc 25
|||||
Db 105 CGATCCTCGTCAGTGCAGTGCAGC 129

RESULT 2
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.

ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
source location/Qualifiers
1..1041
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.00061;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctcgtagtgtagtcagc 25
|||||
Db 105 CGATCCTCGTCAGTGCAGTGCAGC 129

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION B10b gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
AUTHORS Bacteri; Proteobacteria; gamma subdivision; Enterobacteriaceae;
JOURNAL Escherichia.
FEATURES
source location/Qualifiers
1..1084
/organism="Escherichia coli"
/db_xref="taxon:562"
24..1064
/gene="B10b"
24..1064
/gene="B10b"
/codon_start=1
/transl_table=1
/protein_id="CAA00965.1"
/db_xref="GI:490219"
/db_xref="SWISS-PROT:P12996"
/translation="MAHRPMTLSQVLEFEKPLDLLFEAQVHRQHPDPROVOST
LLEIKTGACPEDCRYCPQSSRYKTKGLEAERIMEVEQYLESARAKAGSTRFCMGAAW
KNPIERDMPYLEQWOCVKAMGLACMTLGLTSSQORLANGLDYVNNIDPSPP
YGNITITPTVOERLDLEKVGADGKIVKSGGIVGIGETVVRDAGLLQLANLPTPES
VPIMLIVKVKETPLADNDVDAPFIFTIYAVARIMPTSVYRSAGREQVNEQTAWC
FMAGANSLFTGCKLLITTPNEPEDKDLDFRKLGLNPOOTAVLAGDNEQDQRLDQALMT
PDTEVYNAAL"

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.00061;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctcgtagtgtagtcagc 25
|||||
Db 128 CGATCCTCGTCAGTGCAGTGCAGC 152

RESULT 4
LOCUS E00893 1121 bp DNA
DEFINITION Genomic DNA encoding biotin synthetase.
PAT 29-SEP-1997

ACCESSION E00893
 VERSION E00893.1 GI:2169154
 KEYWORDS JP 1986149091-A/1.
 SOURCE Escherichia coli.
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 1121)
 AUTHORS Hirose, Y., Kojima, T. and Kimura, H.
 TITLE DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND PRODUCTION OF BIOTIN
 JOURNAL Patent: JP 1986149091-A 1 07-JUL-1986;
 NIPPON SODA CO LTD
 COMMENT OS Escherichia coli
 PN JP 1986149091-A/1
 PD 07-JUL-1986
 PF 24-DEC-1984 JP 1984272605
 PI HIROSE YOSHIIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
 C12N1/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
 C12R1:19);
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC *source: strain=Escherichia coli Nsl01;
 CC feature is identified by experimental;
 FH key location/Qualifiers
 FH CDS
 FT 42..1079
 FT /product='biotin synthetase'.
 FEATURES
 source 1..1121
 /organism='Escherichia coli'
 /db_xref='taxon:562'
 BASE COUNT 289 a 296 c 325 g 211 t
 ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 0.00061;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gcatcctgcagtcgagtcagc 25
 ||||||||||||||||||||
 Db 146 CGATCCTGCTCAGTGCAGTGCAGC 170

RESULT 5
 ECOBIO 5793 bp DNA BCT 28-FEB-1994
 LOCUS E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
 DEFINITION (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioc
 protein, and dehydrobiotin synthetase (bioD), complete cds.
 J04423
 J04423 J04423.1 GI:145422
 7,8-diamino-pelargonic acid aminotransferase;
 7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
 bioc gene; bioD gene; bioF gene; biotin synthetase; dehydrobiotin
 synthetase.
 SOURCE Escherichia coli (strain K-12) DNA.
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1 (bases 1 to 5793)
 AUTHORS Otsuka, A.J., Buonocristiani, M.R., Howard, P.K., Plamm, J. and
 Johnson, O.
 TITLE The Escherichia coli biotin biosynthetic enzyme sequences
 JOURNAL J. Biol. Chem. 263, 19577-19585 (1988)
 MEDLINE 89066784
 COMMENT Draft entry and computer-readable sequence [1] kindly submitted by
 A. Otsuka, 09-NOV-1988.
 FEATURES
 Location/Qualifiers

source
 1..5793
 /organism='Escherichia coli'
 /strain='K-12'
 /db_xref='taxon:562'
 complement(98..574)
 /note='ORF 1'
 /codon_start=1
 /transl_table=11
 /protein_id='AAA23513.1'
 /db_xref='GI:455168'
 /translation='MKLISNDLRDCKLPHRYVNGMGVGDGNISPHLAMDVPAGTK
 SEVAVCYDPDAPPTSGMHHVVNLPADRLVLPQGGSLVAMPDGVLTQTRDFCKTG
 YDGAAPPKGETHRYIFTVALDIERIDVDEGSAAMVGFNFHSLASASITAMS'
 complement(633..1925)
 /gene='bioA'
 complement(633..1925)
 /gene='bioA'
 /gene='bioA'
 /codon_start=1
 /transl_table=11
 /product='7,8-diamino-pelargonic acid aminotransferase'
 /protein_id='AAA23514.1'
 /db_xref='GI:457106'
 /translation='MTTDLAFDQRIHWPYTSMTSPLPVYPVVSABGECCELLSGRR
 LVDMGSMWAAIHGVNHPQLNAAMKSQLDAMSHVFGITPADELCRLKLVAMSGRN
 ALECYLADSGVAIVNPMKMALOYQAQGEARORFLFRNGYHGDFTGAMVCDPDN
 SMSLIMKGYLPENLFAPOSRMDGEMDERDVGARLMAHREHIAVIEPIYOGA
 GGRMHPHMLKRIKICDREGILLIADPIATGPGRTKLFCEHAETAPDILGSKA
 LVGGTMTLSATLTITREVAFTISNGEAGCEHGPPTMGKPLACAAANASLIIESDWO
 QOVADIYVQLREQLAPARDAENVADRVLGALGVETHPVNMALQKFEVQGWIR
 PEGLIYIMPPYIILPQQLRLTAAVNRVODETFECQ'
 2012..3052
 /gene='bioB'
 2012..3052
 /gene='bioB'
 /codon_start=1
 /transl_table=11
 /product='biotin synthetase'
 /protein_id='AAA23515.1'
 /db_xref='GI:145425'
 /translation='MAHRPRWTLISOYTELFEPFLDLLEAQVIRQHPDPQVYST
 LLSIKTGACPEDECKYCQPTSRYSKTLGEARLMEVEQVLESARKAKAAGSTRCKMAAM
 KNPHERDMPLYLEQMGVGVKAMGLACMTLGTLSAQORLANAGLDYVNNLDTSPER
 YGNITITRYOERLDITLEKVRDAGIKVCSGIVGTEVKKDRAGLLIQLANLPTPES
 VPINMLVKVKGRLPADNDVDADFIRTAIVGIRIMMPTSYVRLSAREQMNQOTQAMC
 FMAGANSIFYGCKLLTTPNPEBKDLQLRKILGILNQCAVLACGNEDQRLQGLMT
 PDIDETYNAAAL'
 3049..4203
 /gene='bioF'
 3049..4203
 /gene='bioF'
 /codon_start=1
 /transl_table=11
 /product='7-keto-8-amino pelargonic acid synthetase'
 /protein_id='AAA23516.1'
 /db_xref='GI:145426'
 /translation='MSMOEKINAALDARRAALDRRRYPVAGCAGRWLVADDDROYLNF
 SSMDYGLSHHPQIITARMQGAPOEGTISGGSGHVSYSVIOALEELAEGLYSRA
 LLTISPRANOAVIAAMAKEDRIADRLSHSLSEASLSQSOLRRPHNRYTHLAR
 LASPCPGOMGVVTEGVFSMDGDSAPLAEIQQVTOQHNGLWDDAHGTVGEGGGR
 SCMLQKVPPELLAVTPFGKGFVSGAIVLCSIVADYLQFARHLIYSSMPAQAOL
 RASLAVIYRSDEGDAREKLAALITRRAGVODLPTLIADSCSAIOPLIYGDNSRALOL
 AEKLRQGGCVITIRPPTVPAGTARLRLLTLTAHEMODIDRLLEVLHGNG'
 4190..4945
 /gene='bioc'
 /gene='bioc'
 4190..4945
 /gene='bioc'
 /codon_start=1
 /transl_table=11
 /protein_id='AAA23517.1'
 /db_xref='GI:145427'
 /translation='MATVKNQATIAAFGAAGAAHYEHQADLDQROSDALALAMLPQKRYT
 HVLDAAGCGPMKSRHWRERHAQVATLADLSPPMLVQAROKDAADHVLADGIESPLATA
 TFDLAWSNLAVQWGNLSTALRELYRVVBPKVAVFTTLVQGSLLPERIQAMQAVDERP

gene	HANFLPDEIEOISLNGVYOHIILOPITLMEFDALISAMRSLSKIGATHLHGGRDKILL TRSOLOLOLAMPQOOGRYPLTYHLFLGVARE"		
CDS	4938..5597	/gene="biob"	
	4938..5597	/gene="biob"	
	/codon_start=1		
	/transl_table=11		
	/product="dethiobiotin synthetase"		
	/protein_id="AA23518.1"		
	/db_xref="GI:145428"		
	/translation="MSKRYFVGTDTDEVKRTVASCALLQAAKRAGYKRAKYKPVASGS EKRPPEGRNSDALALORNSLSLDYATFADVPYFAEPTSPHITSAGQRPISLWAS; LRAEPQAGLVAGAGGWETPLSDTFADWTOEOLPVLIVGVKLCIGCIHNAITTA OVIOHAGLTLAGVANDVTPPKRHAEMYTTLTRMIPRCWERSPOLQIQMRQPRS T"		
BASE COUNT	1363 a 1554 c 1631 g 1245 t		
ORIGIN	4626 bp upstream of hpaI site; 18 min on K-12 map.		
Query Match	100.0%; Score 25; DB 1; Length 5793;		
Best Local Similarity	100.0%; Pred. No. 0.0005;		
Matches	25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Oy	1	cgactctcgtaagatgcacgtcagc 25	
Db	2116	CGATCTCGTCAGCGTCAGTCACG 2140	
RESULT	6		
LOCUS	A38246	5872 bp DNA	
DEFINITION	Sequence 1 from Patent WO9408023.	PAT	05-MAR-1997
ACCESSION	A38246		
VERSION	A38246.1	GI:2294844	
KEYWORDS			
SOURCE	Escherichia coli.		
ORGANISM	Escherichia coli.		
	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
	Escherichia.		
REFERENCE	1 (bases 1 to 5872)		
AUTHORS	Bitcch,O., Brass,J., Fuhrmann,M. and Shaw,N.		
TITLE	BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN		
JOURNAL	Patent: WO 9408023-A 1 14-APR-1994;		
COMMENT	LONZA AG (CH)		
	Other publication PL 308301 950724		
	Other publication CA 2145400 940414		
	Other publication AU 4820293 940426		
	Other publication HU 71781 960228		
	Other publication SK 42095 951108		
	Other publication CZ 9500809 950913		
	Other publication FI 951547 950331		
	Other publication JP 8501694T 960227.		
FEATURES	Location/Qualifiers		
source	1..5872		
	/organism="Escherichia coli"		
	/strain="DSM498"		
	/db_xref="taxon:562"		
	/clone="PBO30A-15/9"		
	1..96		
promoter	/function="PROMOTER PTAC"		
	/evidence=experimental		
	23..28		
-35_signal	45..49		
-10_signal			
RBS	/standard_name="PROMOTER PTAC"		
	105..119		
	/standard_name="BIOB RBS NO.9"		
	/evidence=experimental		
	117..1157		
gene	/gene="BIOB"		
	117..1157		
CDS	/gene="BIOB"		
	/gene="BIOB"		
	/codon_start=1		
	/transl_table=11		
	/evidence=experimental		
	/product="BIOTIN SYNTHASE"		
	/protein_id="CAA02325.1"		
	/db_xref="GI:2294845"		
	/translation="MAHRRMTLSQVTELEFEKPLDLLPEAOVYHROHFDPRQOVST LSTIKTGACEDCKTCPOSSRTKTELEPRIMEVYOVLESARRKKAAGSTRFGAGAM KNPHERDMPLLEQWGVAKAGLEACMTLGTLSQAOQLANAGLDYYNNHLDTSPEE VGNIIITTKYQERLDLEFKVRDAGIKVCSQGVIGETVKORAGILLQLANLPPSPSS VPJNMLVKVKGTPADIANDVDVDFEITRIIAVRIIMPSTSVYRLSAGEONEQOAC PMAGASIVYSGKILITTPPERDKDLOJFRKIGLNPQOAVIAGNEDQOQLDQALMT PMDVEYNNAAAL"		
RBS	2284..2297		
	/standard_name="BIOC RBS"		
gene	2295..3050		
	/gene="BIOC"		
CDS	2295..3050		
	/gene="BIOC"		
	/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"		
	/codon_start=1		
	/transl_table=11		
	/number=3		
	/product="PROTEIN"		
	/protein_id="CAA02326.1"		
	/db_xref="GI:2294846"		
	/translation="MATVYKKAIAAARGAAHYEHQADLOROSADLLMLPORKYT HYLDACCGRWKSRHMRERHAYVTAIDLSPHLYOARKDAADHTLAGDIESLPLATA TIDLAWSNLAIVOMCGNLSTALRELYRVKPGVAFTLLVQSLPELHQAOAVDERP HANFLPDEIEOISLNGVYOHIILOPITLMEFDALISAMRSLSKIGATHLHGGRPRL TRSOLOLOLAMPQOOGRYPLTYHLFLGVARE"		
RBS	3742..3752		
	/standard_name="BIOA RBS"		
gene	3750..5039		
	/gene="BIOA"		
CDS	3750..5039		
	/gene="BIOA"		
	/standard_name="S-ADENOSYL-L-METHIONINE; 8-AMINO-7-OXONONANATE AMINOTRANSF."		
	/EC_number="2.6.1.62"		
	/codon_start=1		
	/transl_table=11		
	/number=5		
	/evidence=experimental		
	/product="DAPA SYNTHASE"		
	/protein_id="CAA02327.1"		
	/db_xref="GI:2294847"		
	/translation="MTTDDLAPODHIMPTSMTPPYVYVYASACCELIISDGR LVDGSSWMAALHGINHPOLNANKSODIAMSHPFGITHAPAEICRILVMTPOP LHSCVPLADSGVAIVEANRKMALQYQWAGAEKRELFERNYHGDITGASVDDPNS MISLWKGYLPENLFPAPQSRMDGEMDERDVGFAILMAHRIIAVITEPILVQAG GMRMYHEWLKIRIKIDREGILITADEIATGFCRTCKLFCHEAETAPILCIGKAL TSGIMTLSATLTITREVAERTISNGEACGEMHPTPMGNPLCAANASLAILLEGSDML OYADIIEVOLREOLAPARDAEVADRYVIGALIGVETTHPPVMAALQFEVDEGWIRP RGLIILMPRTIILPQOLQRLTAAVNRAVODETFFQO"		
RBS	5088..5100		
	/standard_name="ORFI RBS"		
gene	5098..5574		
	/gene="ORFI"		
CDS	5098..5574		
	/gene="ORFI"		
	/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"		
	/codon_start=1		
	/transl_table=11		
	/number=6		


```

stem_loop      5583..5605
terminator     5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT    1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0005;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 cgatcctcgtcaggtcaggtcagc 25
        |||
Db      221 CGATCCTCGTCAGGTGCGAGTCAGC 245

RESULT 7
LOCUS      A38251      5872 bp      DNA
DEFINITION Sequence 6 from Patent WO9408023.
ACCESSION  A38251
VERSION     A38251.1 GI:2294849
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Escherichia coli.
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL     Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9508089 950913
            Other publication FI 951547 950331
            Other publication JP 85016947 960227.
            Location/Qualifiers
                1..5872
                /organism="Escherichia coli"
                /strain="DSM498"
                /db_xref="taxon:562"
                /clone="PBO30A15-9"
                1141..1156
                /standard_name="BIOF RBS"
                1154..2308
                /gene="BIOF"
                1154..2308
                /gene="BIOF"
                /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                /EC_number="2.3.1.47"
                /codon_start=1
                /transl_table=11
                /number=2
                /evidence=experimental
                /product="KAPA SYNTHASE"
                /protein_id="CAA02329.1"
                /db_xref="GI:2294850"
                /translation="MSWQEKTNALDARADALRRYPVAGGAGRMILVAIDROYLNF
                SSNDYLGISHHPQILIRAWOGCAEOPFGISGSGHVSQSVVHQALEDLEALWLSYSTR
                LFLTSGFANQAVI AAMMAEDRIAADRLSHASLLEASLSPQLRRPANHNVYHLAR
                LLASPCPGCOMVTEGEGVSMGDSAPLAETIQVQOHNGMILVDDAHGTGYIGDQGRG
                SCMLQKVAPELIVTEGKGVSGAAYVCSSTVADYLLQFAHLLIYSTMPADAQAL
                RASLAVIRSDSDGDAERKLAALITFRAGVODLPTLADSCSAIQPLIVGNSALQL
                AKELKQGCWVAITRPVTPAGTARLRLITRAHEMODILLEVHLONG"
RBS
3030..3045
/standard_name="BIOB RBS"
3043..3753
/gene="BIOB"
gene

```

```

CDS
3043..3753
/gene="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYFVYGTDEYGVKTVASCALLQAAKAGYRTAGYKRPVAGSS
EKTEPGRLRSDALALORNSSLQIDAYATVNPYFAEPTSPHIIISAQEGRPISLYMSAG
LRALQOQADWIVVEGAGGMPPLSDPTFEADWVQDEQLPVLLVYGVKLGCIINHMLTA
QVYQIAGULIACVAVANDYTPPKRIAEVMTTLTRIPARLILGELPMLAENENMATK
VYNLAFVDASTGTGTSRL"
BASE COUNT    1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0005;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 cgatcctcgtcaggtcaggtcagc 25
        |||
Db      221 CGATCCTCGTCAGGTGCGAGTCAGC 245

RESULT 8
LOCUS      A93674      5872 bp      DNA
DEFINITION Sequence 1 from Patent EP0798384.
ACCESSION  A93674
VERSION     A93674.1 GI:6741862
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Escherichia coli.
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch,O. and Brass,J.
TITLE      Biotechnological method of producing biotin
JOURNAL     Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES
            source
                1..5872
                /organism="Escherichia coli"
                /strain="DSM498"
                /db_xref="taxon:562"
                /clone="PBO30A-15/9"
                1..96
                /function="PROMOTER PTAC"
                /evidence=experimental
                23..49
                /standard_name="PROMOTER PTAC"
                /evidence=experimental
                105..119
                /standard_name="BIOB RBS NO.9"
                /evidence=experimental
                117..1157
                /gene="BIOB"
                117..1157
                /gene="BIOB"
                /codon_start=1
                /transl_table=11
                /evidence=experimental
                /product="BIOTIN SYNTHASE"
                /protein_id="CAB69590.1"
                /db_xref="GI:6741863"
                /translation="MAHRPRWTLSQVTELEFKRLDLILTEAOQVHRRHPDPPOVYST
                LLSTKTAGCPEDCKYCPQSSRYKTGLAEERLMEVEQVLESARRAKAGSTRFCMGAW

```

KNPHEKMPYLEGMYGVKAMGLEACHTLGLISQAOHLANALDYVNNINLDTSPH
YGNITTTTGYERLDTLKYKADGKVCSSGCVOLGERTKORALLLQALVLPPTPS
VPIINLVKSGPLADNDVDADFOTRTIYAVRIMPISVRLSAGEONNEOQAM:
FMAGNSIFYGCKGLTTPNPEKDLQIFRKLGLNPOTAVLAGDNEQOORLEQALMT
PDTDEYNAAAL"
2284..2297
RBS
/standard_name="BIOC_RBS"
2295..3050
gene
/gene="BIOC"
2295..3050
CDS
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATYKQAIAAFGRAAHYEOHADLQROSDALLAMLPOKYV
HVLGAGCGPMGRHRRHROVATLDSPLMLOAKOKADHYLAGDIESTPLATA
TFDLASMLAYONGCNSTALRELYRYVRKGVATFTLVGSLPELHOQMOAVDEKH
HANRLLPDEIJSUNGHYOHIOPTITLWDDALSAMRSLKIGATVHLEGRDRIIL
TRSLQRLQALAMPQOGRPLTYHLFLGLVARE"
3742..3752
RBS
/standard_name="BIOA_RBS"
3750..5039
gene
/gene="BIOA"
3750..5039
CDS
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONANONATE AMINOTRANSF."/
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDDLAFDORHIMHPYTSPLPYVYVSAACGELLISDCR
LYDGMSSMAAIIHGYNHPOLNAAMKSOIDAMSHYMGITTHAIEICRLVMTPLP
LECYFLADSGSAVEYAMKALOTWQAKGAEKRFETFRNGYHGDITFGASVCDPNS
MHSIMKCYLPENLEAPAPQSRMGEMDERMVGARLMAHRIEIAAVILEPVOGAK
GMRMYHPEWLRKIKICDREGILLADEIATGFRGKLFACAEIAIPIELIGAKI
TGGTMTLSATLITREVAEILISNGEAGCFMGPTFMGNPLACAAANSLALIESGDOU
QVADIEVLEQOLAPARDAEVAADVRLVAGVETTHPYVMAALQKFVEGQVWLRP
FGRLIYMPYILPQOLQRLTAANVANDDETFEFCO"
5088..5100
RBS
/standard_name="ORFI_RBS"
5098..5574
gene
/gene="ORFI"
5098..5574
CDS
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDCKLPHRYFNGYDCDNI SPHLAMDVPAGTK
SPVYTCIDPAPITGSGMMHVVNLPADTRVLPQGFSGSLVAMPDGLQTRDFFGIC
YDGAAPKGTHTHYIFTVHALDIERIDVDEGASGAVGFWHSHLASASTYAMS"
5583..5644
stem_loop
terminator
5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0005;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgatccctcgtcaagctgcagctcagc 25
|||||
Db 221 CGATCCTCGTCAGGTGCGAGTCAC 245
|||||
RESULT 9
A93679
LOCUS A93679 5872 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 6 from Patent EP0798384.
A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (bases 1 to 5872)
Birch, O. and Brass, J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
source location/Qualifiers
1. 5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A15-9"
1141..1156
/standard_name="BIOF_RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONANONATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSWQEKINAALDARRADALRRRYPVAGAGRMV VADROYLNF
LSNLYGSHHQITIRAOQGEBOFGSGGSHYSGT SYVHOLEEELAEMLDYSRA
ILFTGFPANOVIAAMAKEDRIADRLSHASLLEASISPSOLRRRAHNDYHLAR
LLASPCPGQOMVYEGVFSMDGSAPLAEIQVYQHNQGLMVDARHGTGIGQGG
SCMLOKVRBELVYTFKGFVSGAAYICSTVADYLLQFARHLIYSTMPAQAOL
RASLAVIRSDGDAERREKLALITFRFAGVODLPPTLADSCSAIOTPLVGNRQALD
AEKLRQGCQWYAIRPPTVPAGTARLTTLTAHMQDIDRLLEVLNNG"
3030..3045
RBS
/standard_name="BIOB_RBS"
3043..3753
gene
/gene="BIOB"
3043..3753
CDS
/gene="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYEYTGTDTEYKTVASCALLQAKAGTYTACGYKVASGS
EKTEPGLRNSDALALQRRNSLQIDYATVNPYTFAPSPHITSQEGRPTESLVMSAG
LRALEQQAQDWLVLEGAGGMPRLSDTEFTFADWVQEDLPVLLVGVGLGGINHAMLTA
QVIOHAGLTLAGWVANDVTPPGKRHAEMTTLTRMIPABLLGELIPMLAENPENATGK
YINLAFVASTLIGFTSRL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0005;

```

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatctcgtcaggtgcagtcacg 25
    |||||||
Db 221 CGATCCTCGTCAGTCGACGTCAGC 245

RESULT 10
ARI01809
LOCUS ARI01809 5872 bp DNA PAT 14-FEB-2001
DEFINITION Sequence 1 from patent US 6083712.
ACCESSION ARI01809
VERSION ARI01809.1 GI:12812607
KEYWORDS
SOURCE Unknown.
ORGANISM
REFERENCE
1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M., and Shaw, N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
FEATURES
Location/Qualifiers
source
1..5872
/organism="unknown"

BASE COUNT 1318 a 1552 c 1695 g 1307 t

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0005;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatctcgtcaggtgcagtcacg 25
    |||||||
Db 221 CGATCCTCGTCAGTCGACGTCAGC 245

RESULT 11
ARI01810
LOCUS ARI01810 5872 bp DNA PAT 14-FEB-2001
DEFINITION Sequence 6 from patent US 6083712.
ACCESSION ARI01810
VERSION ARI01810.1 GI:12812608
KEYWORDS
SOURCE Unknown.
ORGANISM
REFERENCE
1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M., and Shaw, N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
FEATURES
Location/Qualifiers
source
1..5872
/organism="unknown"

BASE COUNT 1318 a 1552 c 1695 g 1307 t

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0005;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatctcgtcaggtgcagtcacg 25
    |||||||
Db 221 CGATCCTCGTCAGTCGACGTCAGC 245

RESULT 12
AE000180
LOCUS AE000180 11022 bp DNA BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete
genome.
ACCESSION AE000180 U00096

```

```

VERSION AE000180.1 GI:1786398
KEYWORDS
SOURCE Escherichia coli K12.
ORGANISM Escherichia coli K12.
REFERENCE Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
AUTHORS Escherichia.
1 (bases 1 to 11022)
Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
Mau, B., and Shao, Y.
The complete genome sequence of Escherichia coli K-12
Science 277 (5331), 1453-1474 (1997)
JOURNAL MEDLINE
97426617
PUBMED 9278503
REFERENCE
2 (bases 1 to 11022)
Blattner, F.R.
Direct Submission
Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
3 (bases 1 to 11022)
Blattner, F.R.
Direct Submission
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
4 (bases 1 to 11022)
Plunkett, G. III.
Direct Submission
Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using Genemark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
30332 [e-mail: mark@ember.gatech.edu]. Open reading frames that
have been correlated with genetic loci are being annotated with CG
Site Nos., unique ID nos. for the genes in the E. coli Genetic
Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlyn. A public version of the database is accessible
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). ***The E. coli K12 sequence and
its annotations are periodically updated; this is version M55. No
sequence changes. Annotation updates: updated gene identifications
and products: all new functional assignments courtesy of Monica
Riley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'b' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of
labels. This should allow them to be searched for in Entrez as gene
names.

FEATURES
Location/Qualifiers
source
1..11022
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
<1..130
/note="REP (repetitive extragenic palindromic) element;
contains 4 REP sequences"
/gene="ybhC"
/note="b0772"

```


AUTHORS
Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grobeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A. and Blattner, F.R.

TITLE
Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7

JOURNAL
Nature 409 (6819), 529-533 (2001)

MEDLINE
21074935

PUBMED
11206551

REFERENCE
2 (bases 1 to 13501)

AUTHORS
Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grobeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A. and Blattner, F.R.

TITLE
Direct Submission

JOURNAL
Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

FEATURES
Location/Qualifiers
1..13501
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="Enterohemorrhagic"
<1..7576
/note="O-island #36: Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655: Cryptic prophage CP-933K: includes one copy of the 13 bp direct repeat that flanks the prophage"
66..665
/gene="lomK"
/note="20981"
66..665
/gene="lomK"
/function="putative membrane: Other or unknown (Phage or Prophage Related)"
/note="Residues 1 to 199 of 199 are 72.86 pct identical to residues 1 to 199 of 199 from GenPept 118 : g1175327891gb1|AF63231.1|AF151091_2 (AF151091) lom [prophage P-Riba]"
/codon_start=1
/transl_table=1
/product="putative outer membrane protein of prophage CP-933K"
/protein_id="AAG55137.1"
/db_xref="GI:12513752"
/translation="MRKVCATIIISAATCLAVSGVPAWASEHOSNLGAGYLHASTDAPG SDDLNGINVKYRYEFTDTLGLITISFSTANAEDEKTHYSPTRMHEDYVRRKRWNVAG PSYVNVEMFSAVAGAVAYSRTSFSGDYTRVDNKRKTHDVLGSDDAKTSNTSLAW GAGVQNPTESSVADVAYEXSGSGDWRTDGFVGVGRF"
719..2041
/gene="20982"
719..2041
/gene="20982"
/function="putative structure: Structural component (Phage or Prophage Related)"
/note="Residues 164 to 440 of 440 are 68.79 pct identical to residues 381 to 645 of 645 from GenPept 118 : g1145854361gb1|AD25464.1|AF125520.59 (AF125520) putative tail fiber protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=1
/product="putative tail component of prophage CP-933K"
/protein_id="AAG55138.1"
/db_xref="GI:12513753"
/translation="MGMAAVOISGVILKDGAGKPTIONCTIOLKARNSTXYVNTVASE NPDEAGTSMDVEYGGQISVTLVBSGFPSPSHAGTISYVEDSGPTLNDPLFAMIEDDAR PEALRRFQWVEEARHAHEAKKNAGEETISARNAGISASAKAENANADTSADASE SARQAAESAAKSEASASSASAEAAOKSESIQSATDELKSKRTAESAGANAARDA TTTEKARESAESASQASQSRITAAEDAVNRIPTVVGPPGPGPGPAGPOGPKDKGE

gene
CDS
/gene="20985"
2489..3469
/gene="20985"
2489..3469
/gene="20985"
2489..3469
/function="orf: Other or unknown (Phage or Prophage Related)"
/note="Residues 13 to 326 of 326 are 55.73 pct identical to residues 18 to 331 of 336 from GenPept 118 : g1169603671gb1|AF33527.1 (AF170176) hypothetical protein predicted by Glimmer [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55140.1"
/db_xref="GI:12513755"
/translation="MISPIRTTFHNSVNIVSSPCQTFVSFAKEVELKVIDEKPPILE OWFEENPERYKKDEVPIVNTKQHPYLDVNTNARIESDRMIGIPVDCGFVNQKTAFS KLEDEPNVMIITREDVDPSVDRKLSIDYHDIICEORLPTFEDKDEYLVNLLEKELR KLSKQDSLISWYKARNWMPDFRNALALKAGSIFCTYTKNKGISFPGGCIYLD NMILTKIGTITAPDGLISMHYDRKRNDSNIENSAIIVYRSMHPLLEGISFPMHSKVD AHPYDGLKGVKATFNFTPLNHNHFCDFEFNHPNITMNTSQTCSW"
3530..4522
/gene="20986"
3530..4522
/gene="20986"
3530..4522
/function="orf: Other or unknown (Phage or Prophage Related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55141.1"
/db_xref="GI:12513756"
/translation="MKIPSLQSNFNSAPAGYSABIAIPARNAADYVLDIGKRIPL SAADLSNVYESVIRAVHDSRSLIDQHTVDMIGNTVLDLSSRFPRPAAVSGINERK VHTGICIKRYNEELNESSVKLIDDIOSLCTELKEDVGOEPIFPICEAGENDNEPY VSTSVAPKTDSEYEMSGWGLTHEIITHHTGSSDPSGNSFELGPTTEILARVADQELG WSPDKRTGAEBEREAHLRLRNLMALRQAAHHEERKAFERLGTISDRYASDPTF EYSVAVNIQYGFQQHDFPGLAINDLQDANOIOLYHGAPYLFRTGVDDKHMQO"
5350..6231
/gene="20989"
5350..6231
/gene="20989"
5350..6231
/function="orf: Other or unknown (Phage or Prophage Related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55142.1"
/db_xref="GI:12513757"
/translation="MISPYSVNLGCSWNSLTNLTSPDNRLVSSVRAAIVHSNGAOV

```

KGNRTYRVAVTDNKKFCVTRSHSGCFNLLHRLGWKGEISRKTEVLMASPVASAM
ERGIIVSHNRDPLPPVDVAPPELPSVDYNNRLSPGNVKGKNAVYEDADATKVLKM
RTTOSNEEVESEVRCNOYAGSASAKITGNNDIIGIMDKINBSLINSISLIVU
AEHAIDYDPLREKQKILFVDFTFTNLYBRAKNEFNPIDISSINVDRSWSQIMU
SYHGKODLISVLSKI"
6462..7160
/ gene="20990"
6462..7160
/ gene="20990"
/ function="orf; Other or unknown (Phage or Prophage
Related)"
/ note="No significant matches"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AG55143.1"
/db_xref="GI:12513758"
/translation="MRPTSLNLYHOSRSSSDPDIESIAKSSVQWIKNNPOLRF
OGTDHNIYQOIEALDKIGSETGKRVLLNIESIRLSTETVYIHLSSRUGMAIHD
IDAENHNGTSDPHCNNAVEYPCGEGISVDEHATTVEHLHVFHNLNGERLKVES
SRAESOKYSPILLEEARVGLGAFSEEVLSENKFERIEGMPRTSYPSXSALIHDNI
VSLGEQVRLHPL"
/ complement(7668..8144)
/ gene="ybhb"
/ note="20992"
/ complement(7668..8144)
/ gene="ybhb"
/ function="orf; Unknown function"
/ note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655: 80773"
/codon_start=1
/transl_table=1
/product="orf; hypothetical protein"
/protein_id="AAG5144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDCKLPHRFHNGYDGNISPHLAMDVPAGTK
SFVATCYDPAPTGSGMMHVVNLPADRTLPQFGSGIYVAMPDGLQTRDFEGTG
YDGAAPKGETHYRIFTVAIDVERIDVDEGASGAVGFWHSHLASASTAMFS"
/ complement(8203..9492)
/ gene="bioda"
/ note="20993"
/ complement(8203..9492)

Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 0.00045;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgatctcgtcaagtcgaagtcacg 25
|||||
db 9683 CGATCTCTGTCAGCTGCGAGTCAGC 9707

RESULT 14
AP002553 297816 bp DNA BCT 07-MAR-2001
LOCUS Escherichia coli O157:H7 DNA, complete genome, section 4/20.
DEFINITION AP002553 BA000007
VERSION AP002553.1 GI:13360211
KEYWORDS
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (sites)
Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C. and Shinagawa,H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak

JOURNAL GENES Genet. Syst. 74 (5), 227-239 (1999)
MEDLINE 20198780
REFERENCE
2 (sites)
Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
JOURNAL 20573356
MEDLINE
REFERENCE
3 (sites)
Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
Shinagawa,H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
JOURNAL 20564182
MEDLINE
REFERENCE
4 (sites)
Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shinagawa,H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
JOURNAL 21156231
MEDLINE
REFERENCE
5 (bases 1 to 297816)
Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
Hayashi,T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail:Ken@gen-info.osaka-u.ac.jp,
URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
Fax:81-6-6879-2047)
genome project.
COMMENT Location/Qualifiers
FEATURES
SOURCE
1..297816
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
79..1245
/ gene="ECs0753"
79..1245
/ gene="ECs0753"
/ note="ECs0753"
100 in 388 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=1
/translation="MNLHEYQAKOLFARYGLPAPVGYACTTPREAEFASKIGAPWY
VKCOVHAGRGKAGGVVYNSKEDIKFAFEKWLKRLVYOTDANGOPVNOILVEAT
DIAKEIYLGAEKLVQDFTKIFNGIATIFLERDLALIEINPLVITKQDILCDGKGA
DGNALFROPDLRENRDQSDPRAQAQWELNVALDNGVNGVNGAGLAWGTMDIV
KLHGGEPAFLDVGGAFTKERYEAKIILISDKVAVNLVNIIFGIVKCDLIADGIIG
AAVEGVNVPVYVLEGNNAELGAKKILADSLNIIAAKGLTDAQGVAAVEK"
1245..2114
/ gene="ECs0754"
1245..2114
/ gene="ECs0754"
/ note="ECs0754"
/ note="similar to SUCD_ECOLI g11786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1

```

	gene	DIAAKATIKARLDSEPMPEVLEKHHIVYVGPAPKPBENMAALSLPTTGGRDXYIDPFIQ QAAGSLVLSLTKSGNRSDQYDADACHRRHGGNSISGGAALLAEOEVVNSLRCLFEPBLG MEAYMMHEVENLPFFILVDKGNFFSOFEGQHRCASCPAH"
CDS	gene	/complement(6092..7372) /gene="ECs0758" /complement(6092..7372) /note="probable transport protein,similar to glutamate/aspartate transport proteins (proton glutamate symport proteins) e.g. [Bacillus stearothermophilus] gII12167[IsipIP24943]GLMT_BACST percent identity 38 in 416 aa,also similar to C4-dicarboxylate transporter" /codon_start=1 /transl_table=1 /evidence=not_experimental /product="putative transport protein" /protein_id="BAB34181.1" /db_xref="GI:13360217"
CDS	gene	/translation="MKRTSIITTMILALVLGVLMNTPETAKLYAOEISITFT TIFRLIKMIIAPLVSTLVVGIAKKGDKAAGRITSKTLFLTCASLISAIGLITVV NEFMFGGINFVAHGAEITGTGVNAEPTLKVTISHAFPSVIDAMAHNEIIOIVFSF FLGSLSTAIGKSAGSAIVHADSLAHMLKTGYVMFAFLVFAMISALIIERGIAVMS VSAGIPMGEEFPFLIMLVILIGLAIIVYGPCIRLTRIASPALTAFTSSSAAPP CTLKLKDPOGSPKIASFVPIGTSYNLGSNAFCFAYPIAQCNHLHSIGDTIPP LLILMLSKGAGVPRAKSNVVIATINQENTPEAGLITLMGDPFLDMGRSATWMSM AMGAAMSWECEHFEGCGCKAKIKPNESVALP" /complement(7533..7850) /gene="ECs0759" /complement(7533..7850) /note="unknown" /codon_start=1 /transl_table=1 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB34182.1"
Query Match		100.0%; Score 25; DB 1; Length 297816; Best Local Similarity 100.0%; Pred. No. 0.00031;
Matches	25; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1	cgaatcctgcaggtgcaagtcagc 25
Db	92415	CGATCCTCGTCGACGTCAGTCAGC 92439
RefSeq	15	SMABIO
LOCUS	SMABIO	7215 bp DNA BCT 04-FEB-1999
DEFINITION	S. marcescens biotin operon, complete sequence.	
ACCESSION	D17468	
VERSION	D17468.1	GI:402530
KEYWORDS	7,8-diamino-pelargonic acid aminotransferase; 7-keto-8-amino-pelargonic acid synthetase; DAPA aminotransferase; DTP synthetase; KAPA synthetase; biotin operon; biotin synthetase; dethiolobiotin synthetase. Serratia marcescens (strain:sf41) DNA.	
SOURCE	Serratia marcescens	
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Serratia.	
REFERENCE	1 (bases 1 to 7215)	
AUTHORS	Sakurai,N.	
TITLE	Direct Submission	
JOURNAL	Submitted (25-AUG-1993) to the DDBJ/EMBL/GenBank databases. Naoki Sakurai, Tanabe Seiyaku Co., Ltd., Res Lab of Applied Biochemistry; 2-50, Kawagishi-2-chome, Toda-shi, Saitama 335, Japan (E-mail:GHC01101@niftyserve.or.jp, nsakurai@dcbj.nig.ac.jp, Tel.:048-433-2545, Fax:048-433-2540)	
REFERENCE	2 (bases 1 to 7215)	
AUTHORS	Sakurai,N., Imai,Y., Akatsuka,H., Kawai,E., Komatsu bara,S. and Tosa,T.	
TITLE	Complete nucleotide sequence of biotin operon of Serratia marcescens	

JOURNAL COMMENT

Unpublished (1993)
Submitted (25-Aug-1993) to DDBJ by:
Naoki Sakurai
Res. Lab. of Applied Biochemistry
Tanabe Seiyaku Co., Ltd.
2-50 Kawagishi-2-chome
Toda, Saitama 335
Japan
Phone: 048-433-2545
Email: nsakurai@dabj.nig.ac.jp
Fax: 048-433-2540.

FEATURES

source

Location/Qualifiers
1. 7215

/organism="Serratia marcescens"

/strain="Sr41"

/db_xref="taxon:615"

/complement(1353..2630)

/gene="bioc"

/complement(1353..2630)

/gene="bioc"

/EC_number="2.6.1.62"

/codon_start=1

/transl_table=1

/product="7,8-diamino-pelargonic acid aminotransferase"

(DAPA aminotransferase)"

/protein_id="BA04284.1"

/db_xref="GI:402531"

/translation="MSVYASDLAFDQRIHWPYTSMSRPLPCYPYESAGVELQLADG
RSLVDGSSWMAAIIHGYNHPRINQASQLEKSHVMFGITHPAALSICRIIVEMIP
EALQGVFLADSGVAEVEVSLKMALOYMARGERFORILTLRHGYHGDTPGMSVCD
NSHSLYGYLAPLPLFATAPQCRDEDEMRDILAPFALLEHNAAGEVAATLPPVOC
AGSMRIYPTLYKRDVAVATIKLLIADENIAGKRTGKLEFACAGVAPPLICG
KALTGYMTLSATITTHVAETISNGAAGCFMPTMCPPLACAVADAALAEKR
WQAVNSAIFETOLKRELLPLGRCRPHRVLGATGVEMREPVDAVLEQGEVGEVW
IRPGKLIYLMPPYIIIEQLSRLTAVAADAR"

2716..3756

/gene="bioc"

2716..3756

/gene="bioc"

/codon_start=1

/transl_table=1

/product="biotin synthetase"

/protein_id="BA04285.1"

/db_xref="GI:402532"

/translation="MMADRIHMTYGOALFEDKPLLELLFEAGYVHHQHEDPPQVVS
TLSTIKGACPEDECKYCPOSPRYKTCLESELDVQVEVLESARAKANSTTRCMGA
WKNPHERDMPFLQOVGVKAMGETMTGLTGDQAEKLAEGLDVYHNLDTSP
EYGSITTRRSQERLADLVKRVAGIKVCGGIVGLGETVRAGLVLQALNPKPE
SVPIMLVKKGTPLADNDVDPPDFITRTAVAIMPSSYVRLSAGREMDQOTAM
CFMAGANSIFYGKLTTPNFEEDKDLQLRKILGLNPQATVTEHGMDQOOVLAQQL
NADPAEFYNAAP"

3756..4904

/gene="bioc"

3756..4904

/gene="bioc"

/codon_start=1

/transl_table=1

/product="7-keto-8-aminopelargonic acid synthetase (KAPA
synthetase)"

/protein_id="BA04286.1"

/db_xref="GI:402533"

/translation="MSWQRIEQLAERLNAAYRRQTEGGNGROIIGLDLYLNF
SGNDYLSODARVITAMOGGARYGVNGSGSVTGSAAHOLEBOLALMGYPRAL
LPISYANQAVIALMOKCDRIADRLSHASLIEAAGAVGRAPPAOSTAGLARI
LAKCDGQORLAIVTGLFSGMGDAPLAEHLRTRACAGMLVDDAHGIGRGGKRS
CQGVGPELLVATFEKAFVSGAIVCDATAEYLLQFARHLIYSTAMPACALQ
AALARIGDGLRRLQDNIRRFROGAPLALTLTSDITAIQPLVGDNRADLALTR
LREGLWVSAIRPPTVPBGARLRLTLTAHOSODIDRLLEVLNDSO"

4888..5655

/gene="bioc"

4888..5655

/gene="bioc"

/function="synthesis of pimeloyl-CoA"

FEATURES

source

1..7318

/organism="Mus musculus"

/strain="C57BL6"

/db_xref="taxon:10090"

/clone_lib="Lambda ziploxx"

/dev_stage="adult"

BASE COUNT

1347 a 2225 c 2367 g 1276 t

ORIGIN

/codon_start=1
/transl_table=11
/product="the product of bioc"
/protein_id="BA04287.1"
/db_xref="GI:402534"
/translation="MFSANDYNNKQAVASAFSRAAGSYDAAMALORDVGPRLMGSS
HPGEBLLDAGCGTGFSSRMKREKRYVALDLARMDLVAKORARAHYLLGLDTEQY
LPDAAMDICFSSLVYQWCSDLPALAEILYRTYRREGVITLSTLAGSLQELGDAMQV
DEBRVNAFLPLQIRIACAVRHELVLELRLNPDVMTMRSLKLGIGATYHLOGRE
GGLMSRGRIALAOAAVPCROGQFPLSYHLAVGYIRE"
5648..6331
/gene="bioc"
5648..6331
/gene="bioc"
/function="conversion from 7,8-diaminopelargonic acid into
dethiobiotin"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase (DTB synthetase)"
/protein_id="BA04288.1"
/db_xref="GI:912453"
/translation="MSKRFVYGTDEVEKSVASSALLQANRAGYSPAGYPAVSGS
EMTAEGLRNGDALQANSVVALDYDEVNYPVFAEPTSPHIVSADDEGRIDARLSDG
LRRLBORADWVLEVGAGWEPPLSAEYFADWVROELPVLYGIRKIGCINHAIVTA
QAVQAGLITLACWIANVDVGAGAPRRHDEYLTALTRMLPRRCMAKSRTRCRRERAPLQ
YIDISLILAO"

Query Match

Best Local Similarity 72.0%; Score 18; DB 1; Length 7215;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cytcaggtgcaggtcagc 25

|||||

Db 2830 CGTCAGGTGCAGGTACG 2847

RESULT 16

AB026497/c

LOCUS

DEFINITION

AB026497 7318 bp mRNA

AB026497

AB026497.1 GI:7416031

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

3539 4327: contig of 789 bp in length
4328 4427: gap of 100 bp
4428 5234: contig of 807 bp in length
5235 5334: gap of 100 bp
5335 6129: contig of 795 bp in length
6130 6229: gap of 100 bp
6230 7040: contig of 811 bp in length
7041 7140: gap of 100 bp
7141 7931: contig of 791 bp in length
7932 8031: gap of 100 bp
8032 8810: contig of 779 bp in length
8811 8910: gap of 100 bp
8911 9696: contig of 786 bp in length
9697 9796: gap of 100 bp
9797 10596: contig of 800 bp in length
10597 10696: gap of 100 bp
10697 11493: contig of 797 bp in length
11494 11593: gap of 100 bp
11594 12399: contig of 806 bp in length
12400 12499: gap of 100 bp
12500 13296: contig of 787 bp in length
13297 13396: gap of 100 bp
13397 14152: contig of 756 bp in length
14153 14252: gap of 100 bp
14253 15034: contig of 782 bp in length
15035 15134: gap of 100 bp
15135 15922: contig of 788 bp in length
15923 16022: gap of 100 bp
16023 16795: contig of 773 bp in length
16796 16895: gap of 100 bp
16896 17698: contig of 803 bp in length
17699 17798: gap of 100 bp
17799 18568: contig of 770 bp in length
18569 18668: gap of 100 bp
18669 19457: contig of 789 bp in length
19458 19557: gap of 100 bp
19558 20342: contig of 785 bp in length
20343 20442: gap of 100 bp
20443 21239: contig of 797 bp in length
21240 21339: gap of 100 bp
21340 22122: contig of 783 bp in length
22123 22222: gap of 100 bp
22223 22995: contig of 773 bp in length
22996 23095: gap of 100 bp
23096 23836: contig of 741 bp in length
23837 23936: gap of 100 bp
23937 24720: contig of 784 bp in length
24721 24820: gap of 100 bp
24821 25590: contig of 770 bp in length
25591 25690: gap of 100 bp
25691 26480: contig of 790 bp in length
26481 26580: gap of 100 bp
26581 27368: contig of 788 bp in length
27369 27468: gap of 100 bp
27469 28258: contig of 790 bp in length
28259 28358: gap of 100 bp
28359 29134: contig of 776 bp in length
29135 29234: gap of 100 bp
29235 30016: contig of 782 bp in length
30017 30116: gap of 100 bp
30117 30897: contig of 781 bp in length
30898 30997: gap of 100 bp
30998 31770: contig of 773 bp in length
31771 31870: gap of 100 bp
31871 32662: contig of 792 bp in length
32663 32762: gap of 100 bp
32763 33558: contig of 796 bp in length
33559 33658: gap of 100 bp
33659 34448: contig of 790 bp in length
34449 34548: gap of 100 bp
34549 35346: contig of 798 bp in length
35347 35446: gap of 100 bp
35447 36251: contig of 805 bp in length

36252 36351: gap of 100 bp
36352 37136: contig of 785 bp in length
37137 37236: gap of 100 bp
37237 38021: contig of 785 bp in length
38022 38121: gap of 100 bp
38122 38908: contig of 787 bp in length
38909 39008: gap of 100 bp
39009 39800: contig of 792 bp in length
39801 39900: gap of 100 bp
39901 40683: contig of 783 bp in length
40684 40783: gap of 100 bp
40784 41551: contig of 768 bp in length
41552 41651: gap of 100 bp
41652 42409: contig of 758 bp in length
42410 42509: gap of 100 bp
42510 43302: contig of 793 bp in length
43303 43402: gap of 100 bp
43403 44204: contig of 802 bp in length
44205 44304: gap of 100 bp
44305 45097: contig of 793 bp in length
45098 45197: gap of 100 bp
45198 45976: contig of 779 bp in length
45977 46076: gap of 100 bp
46077 46856: contig of 780 bp in length
46857 46956: gap of 100 bp
46957 47735: contig of 779 bp in length
47736 47835: gap of 100 bp
47836 48609: contig of 774 bp in length
48610 48709: gap of 100 bp
48710 49519: contig of 810 bp in length
49520 49619: gap of 100 bp
49620 50405: contig of 786 bp in length
50406 50505: gap of 100 bp
50506 51282: contig of 777 bp in length
51283 51382: gap of 100 bp
51383 52152: contig of 770 bp in length
52153 52252: gap of 100 bp
52253 53044: contig of 792 bp in length
53045 53144: gap of 100 bp
53145 53941: contig of 797 bp in length
53942 54041: gap of 100 bp
54042 54836: contig of 795 bp in length
54837 54936: gap of 100 bp
54937 55730: contig of 794 bp in length
55731 55830: gap of 100 bp
55831 56616: contig of 786 bp in length
56617 56716: gap of 100 bp
56717 57498: contig of 782 bp in length
57499 57598: gap of 100 bp
57599 58381: contig of 783 bp in length
58382 58481: gap of 100 bp
58482 59250: contig of 769 bp in length
59251 59350: gap of 100 bp
59351 60139: contig of 789 bp in length
60140 60239: gap of 100 bp
60240 61030: contig of 791 bp in length
61031 61130: gap of 100 bp
61131 61917: contig of 787 bp in length
61918 62017: gap of 100 bp
62018 62811: contig of 794 bp in length

Query Match 68.0%: Score 17; DB 2; Length 69900;

Best Local Similarity 100.0%: Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cgtcaggtcaggtcag 24
|||||

Db 19825 CGTCAGGTGAGGTGAG 19809

RESULT 18

AC069132 151340 bp DNA HTG 17-JUN-2000
LOCUS

DEFINITION Homo sapiens chromosome 5 clone RP11-556i8, *** SEQUENCING IN
PROGRESS ****, 62 unordered pieces.
ACCESSION AC069132 GI:8469032
VERSION AC069132.2
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 151340)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 151340)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (18-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jun 12, 2000 this sequence version replaced gi:7924001.
----- Genome Center -----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1502: contig of 1502 bp in length
* 1503 1602: gap of unknown length
* 1603 2741: contig of 1139 bp in length
* 2742 2841: gap of unknown length
* 2842 4036: contig of 1195 bp in length
* 4037 4136: gap of unknown length
* 4137 5620: contig of 1484 bp in length
* 5621 5720: gap of unknown length
* 5721 7080: contig of 1360 bp in length
* 7081 7180: gap of unknown length
* 7181 8839: contig of 1659 bp in length
* 8840 8939: gap of unknown length
* 8940 9987: contig of 1048 bp in length
* 9988 10087: gap of unknown length
* 10088 11156: contig of 1069 bp in length
* 11157 11256: gap of unknown length
* 11257 12404: contig of 1148 bp in length
* 12405 12504: gap of unknown length
* 12505 13861: contig of 1357 bp in length
* 13862 13961: gap of unknown length
* 13962 15241: contig of 1280 bp in length
* 15242 15341: gap of unknown length
* 15342 17009: contig of 1668 bp in length
* 17010 17109: gap of unknown length
* 17110 18897: contig of 1788 bp in length
* 18898 18997: gap of unknown length
* 18999 20389: contig of 1392 bp in length
* 20390 20489: gap of unknown length
* 20490 22042: contig of 1553 bp in length
* 22043 22142: gap of unknown length
* 22143 24057: contig of 1915 bp in length
* 24058 24157: gap of unknown length
* 24159 26189: contig of 2032 bp in length
* 26190 26289: gap of unknown length
* 26290 27959: contig of 1670 bp in length
* 27960 28059: gap of unknown length
* 28060 30067: contig of 2008 bp in length
* 30068 30167: gap of unknown length
* 30168 32456: contig of 2289 bp in length
* 32457 32557: gap of unknown length
* 32558 34597: contig of 2041 bp in length
* 34598 34697: gap of unknown length
* 34698 36611: contig of 1914 bp in length

36612 36711: gap of unknown length
* 36712 38677: contig of 1966 bp in length
* 38678 38777: gap of unknown length
* 38778 40974: contig of 2197 bp in length
* 40975 41074: gap of unknown length
* 41075 42667: contig of 1592 bp in length
* 42668 42766: gap of unknown length
* 42767 44292: contig of 1526 bp in length
* 44293 44392: gap of unknown length
* 44393 46032: contig of 1640 bp in length
* 46033 46133: gap of unknown length
* 46133 48815: contig of 2683 bp in length
* 48816 48915: gap of unknown length
* 48916 51058: contig of 2143 bp in length
* 51059 51158: gap of unknown length
* 51159 53094: contig of 1936 bp in length
* 53095 53194: gap of unknown length
* 53195 54724: contig of 1530 bp in length
* 54725 54824: gap of unknown length
* 54825 57019: contig of 2195 bp in length
* 57020 57119: gap of unknown length
* 57120 59736: contig of 2617 bp in length
* 59737 59836: gap of unknown length
* 59837 61510: contig of 1674 bp in length
* 61511 61610: gap of unknown length
* 61611 63247: contig of 1637 bp in length
* 63248 63347: gap of unknown length
* 63348 65262: contig of 1915 bp in length
* 65263 65362: gap of unknown length
* 65363 67493: contig of 2131 bp in length
* 67494 67593: gap of unknown length
* 67594 70740: contig of 3147 bp in length
* 70741 70840: gap of unknown length
* 70841 72256: contig of 1416 bp in length
* 72257 72356: gap of unknown length
* 72357 73456: contig of 3100 bp in length
* 73457 75556: gap of unknown length
* 75557 77328: contig of 1772 bp in length
* 77329 77428: gap of unknown length
* 77429 79851: contig of 2423 bp in length
* 79852 79951: gap of unknown length
* 79952 82877: contig of 2926 bp in length
* 82878 82977: gap of unknown length
* 82978 84859: contig of 1882 bp in length
* 84860 87856: contig of 2897 bp in length
* 87857 87956: gap of unknown length
* 87957 91677: contig of 3721 bp in length
* 91678 91777: gap of unknown length
* 91778 94536: contig of 2759 bp in length
* 94537 94637: gap of unknown length
* 94638 97377: contig of 2741 bp in length
* 97378 97477: gap of unknown length
* 97478 100981: contig of 3504 bp in length
* 100982 101081: gap of unknown length
* 101082 103958: contig of 2877 bp in length
* 103959 104058: gap of unknown length
* 104060 106903: contig of 2845 bp in length
* 106904 107003: gap of unknown length
* 107004 109662: contig of 2659 bp in length
* 109663 109762: gap of unknown length
* 109763 113672: contig of 3910 bp in length
* 113673 113772: gap of unknown length
* 113773 118353: contig of 4581 bp in length
* 118354 118453: gap of unknown length
* 118454 121541: contig of 3088 bp in length
* 121542 121641: gap of unknown length
* 121642 124122: contig of 2481 bp in length
* 124123 124222: gap of unknown length
* 124223 128114: contig of 3892 bp in length
* 128115 128214: gap of unknown length
* 128215 131684: contig of 3470 bp in length
* 131685 131784: gap of unknown length


```

misc_feature 114829..125622
                /note="assembly-fragment"
misc_feature 125723..156844
                /note="assembly-fragment"
misc_feature 156945..161742
                /note="assembly-fragment"
                clone_end:r7
                vector_side:right"
BASE COUNT 45576 a 33084 c 34118 g 47963 t 1001 others
ORIGIN
Query Match 68.0% Score 17: DB 2: Length 161742:
Best Local Similarity 100.0% Pred No. 9.9;
Matches 17: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 8 cgtcaggtcagtcagtcag 24
    |||
Db 70155 CGTCAGGTGACGTCAG 70139

RESULT 20
AC041048/c 168360 bp DNA HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 5 clone RP11-759J14, WORKING DRAFT
DEFINITION AC041048.3 GI:7770701
VERSION AC041048.3 GI:7770701
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            Walerston, R.H.
            The sequence of Homo sapiens clone
            2 (bases 1 to 168360)
            Unpublished
            Waterston, R.H.
            Direct Submission
            Submitted (11-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            On May 12, 2000 this sequence version replaced gi:7622530.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0759J14
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 157456 bases at least Q40
Consensus quality: 161145 bases at least Q30
Consensus quality: 163068 bases at least Q20
Insert size: 173000; agarose-1p
Insert size: 166760; sum-of-ctrls
Quality coverage: 4.08 in Q20 bases; sum-of-ctrls
Quality coverage: 4.27 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

FEATURES
    source
        1..1711 contig of 1711 bp in length
        1712..1811 gap of unknown length
        1812..6135 contig of 4324 bp in length
        6136..6236 gap of unknown length
        6236..8687 contig of 2452 bp in length
        8688..11647 gap of unknown length
        11648..11747 contig of 2860 bp in length
        11748..13718 gap of unknown length
        13719..13819 gap of unknown length
        13819..18016 contig of 4198 bp in length
        18017..18117 gap of unknown length
        18117..21570 contig of 3454 bp in length
        21571..21671 gap of unknown length
        21671..24723 contig of 3053 bp in length
        24724..24823 gap of unknown length
        24824..32650 contig of 7827 bp in length
        32651..32750 gap of unknown length
        32751..32751 contig of 13741 bp in length
        32751..46491 gap of unknown length
        46492..46591 gap of unknown length
        46592..60127 contig of 13536 bp in length
        60128..60227 gap of unknown length
        60228..74481 contig of 14254 bp in length
        74482..74581 gap of unknown length
        74582..90886 contig of 16505 bp in length
        90887..90887 gap of unknown length
        90887..104984 contig of 13998 bp in length
        104985..105084 gap of unknown length
        105085..127380 contig of 22295 bp in length
        127380..127479 gap of unknown length
        127480..147028 contig of 19549 bp in length
        147029..147129 gap of unknown length
        147129..168360 contig of 21232 bp in length.
    Location/Qualifiers
        1..168360
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="5"
        /clone="RP11-759J14"
        1..1711
        /note="assembly_name:Contig6"
        1812..6135
        /note="assembly_name:Contig7"
        6236..8687
        /note="assembly_name:Contig8"
        clone_end:SP6
        vector_side:right"
        8788..11647
        /note="assembly_name:Contig9"
        11748..13718
        /note="assembly_name:Contig10"
        13819..18016
        /note="assembly_name:Contig11"
        18117..21570
        /note="assembly_name:Contig12"
        21671..24723
        /note="assembly_name:Contig13"
        24824..32650
        /note="assembly_name:Contig14"
        32751..46491
        /note="assembly_name:Contig15"
        46592..60127
        /note="assembly_name:Contig16"
        60228..74481
        /note="assembly_name:Contig17"
        74582..90886
        /note="assembly_name:Contig18"
        90987..104984
        /note="assembly_name:Contig19"
        105085..127379
        /note="assembly_name:Contig20"
        127480..147028
        /note="assembly_name:Contig21"

```

misc-feature 147129, .168360
 /note="assembly_name:Contig22"
 BASE COUNT 48985 a 34329 c 34643 g 48785 t 1618 others
 ORIGIN

Query Match 68.0%; Score 17; DB 2; Length 168360;
 Best Local Similarity 100.0%; Pred. No. 9.9;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 8 cgtcagtcagtcag 24
 |||||||
 Db 153451 CGTCAGTCAGTCAG 153435

RESULT 21
 AC060833/c
 LOCUS Homo sapiens chromosome 4 clone RP11-556F3, WORKING DRAFT SEQUENCE,
 DEFINITION 4 unordered pieces.
 AC060833
 VERSION AC060833.3 GI:9838272
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 176379)
 TITLE The sequence of Homo sapiens clone
 REFERENCE 2 (bases 1 to 176379)
 AUTHORS Waterston, R.H.
 JOURNAL Unpublished
 TITLE Direct Submission
 AUTHORS Submitted (20-APR-2000) Genome Sequencing Center, Washington
 JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 COMMENT On Aug 17, 2000 this sequence version replaced gi:8980022.

Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 Project Information
 Center project name: H.NH0556F03
 Summary Statistics
 Sequencing vector: M13; 100%
 Chemistry: Dye-Primer ET; 100% of reads
 Assembly: Dye-terminator Big Dye; 0% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 173730 bases at least Q40
 Consensus quality: 174579 bases at least Q30
 Consensus quality: 175046 bases at least Q20
 Insert size: 174000; agarose-ftp
 Insert size: 177415; sum-of-coverage
 Quality coverage: 6.55 in Q20 bases; sum-of-coverage
 Quality coverage: 6.55 in Q20 bases; sum-of-coverage
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1419: contig of 1419 bp in length
 * 1420 1519: gap of unknown length
 * 1520 10016: contig of 8497 bp in length
 * 10017 10116: gap of unknown length
 * 10117 79219: contig of 69103 bp in length
 * 79220 79319: gap of unknown length

FEATURES * 79320 176379: contig of 97060 bp in length.
 Location/Qualifiers
 source

misc-feature
 1. 1419
 /note="assembly_name:Contig5"
 1520. 10016
 /note="assembly_name:Contig6
 clone_end:SP6
 vector_side:right"
 10117. 79219
 /note="assembly_name:Contig7"
 79320. 176379
 /note="assembly_name:Contig8
 clone_end:17
 vector_side:left"
 misc-feature
 misc-feature
 BASE COUNT 49569 a 33984 c 35497 g 57025 t 304 others
 ORIGIN

Query Match 68.0%; Score 17; DB 2; Length 176379;
 Best Local Similarity 100.0%; Pred. No. 9.8;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 8 cgtcagtcagtcag 24
 |||||||
 Db 81954 CGTCAGTCAGTCAG 81938

RESULT 22
 AC022781/c
 LOCUS Mus musculus chromosome 11 clone RP23-199H17 map 11, WORKING DRAFT
 DEFINITION SEQUENCE, 8 unordered pieces.
 AC022781.6 GI:14595917
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 1 (bases 1 to 220166)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Beckert, R., Bede, F.,
 Bonuskavsky, L., Bouckgalter, B., Brown, A., Burkett, G., Castle, A.,
 Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
 Dearrellano, K., Dewar, K., Domino, M., Doyle, M., Fennestor, J.,
 Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
 Gardyna, S., Grant, G., Hagos, B., Heatford, A., Horton, L.,
 Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,
 MacDonald, P., Margulis, N., McEwan, P., McGurk, A., McKernan, K.,
 McPheters, R., Meldrim, J., Menus, L., Morrow, J., Naylor, J.,
 Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
 Plette, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
 Tirelli, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
 Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 4, 2001 this sequence version replaced gi:10305259.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 199_H.17
----- Summary Statistics
Sequencing vector: M13: M77815; 43% of reads
Sequencing vector: Plasmid: n/a; 57% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 217289 bases at least Q40
Consensus quality: 218321 bases at least Q30
Consensus quality: 218800 bases at least Q20
Insert size: 206000; agarose-fp
Insert size: 219466; sum-of-contigs
Quality coverage: 9.5 in Q20 bases; agarose-fp
Quality coverage: 8.9 in Q20 b.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 22567: contig of 22567 bp in length
* 22568 22667: gap of 100 bp
* 22668 28303: contig of 5636 bp in length
* 28304 28403: gap of 100 bp
* 28404 109561: contig of 81158 bp in length
* 109562 109661: gap of 100 bp
* 109662 118270: contig of 8609 bp in length
* 118271 118370: gap of 100 bp
* 118371 140406: contig of 22036 bp in length
* 140407 140506: gap of 100 bp
* 140507 158389: contig of 17883 bp in length
* 158390 158489: gap of 100 bp
* 158490 211583: contig of 53094 bp in length
* 211584 211683: gap of 100 bp
* 211684 220166: contig of 8483 bp in length.
FEATURES
source
1. 220166
   /organism="Mus musculus"
   /db_xref="taxon:10090"
   /chromosome="11"
   /map="11"
   /clone="RP23-199H17"
   /clone_lib="RPC1-23 Female Mouse BAC"
1. 22567
   /note="assembly-fragment"
   /clone_end:SP6
   /vector_side:left"
misc_feature
22668..28303
   /note="assembly-fragment"
28404..109561
   /note="assembly-fragment"
109662..118270
   /note="assembly-fragment"
118371..140406
   /note="assembly-fragment"
140507..158389
   /note="assembly-fragment"
158490..211583
   /note="assembly-fragment"
211684..220166
   /note="assembly-fragment"
   /clone_end:77
   /vector_side:right"
misc_feature
53983 a 54774 c 55714 g 54982 t 713 others
ORIGIN

```

```

Query Match 8 8 cgcacgctgcacgctcac 24
Best Local Similarity 100.0%; Score 17; DB 2; Length 220166;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 37316 CGTCAGCTGCACGCTCAC 37300
-----
RESULT 23
AL591065
LOCUS
DEFINITION
MUS musculus chromosome 11 clone RP23-199H17, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION
AL591065.2 GI:13990285
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 22212)
REFERENCE
Plumb,B.
AUTHORS
Direct Submission
TITLE
Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
JOURNAL
On May 7, 2001 this sequence version replaced gi:13938846.
COMMENT
-----
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
-----
Project Information
Center project name: BM199H17
-----
Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 218929 bases at least Q40
Consensus quality: 219403 bases at least Q30
Consensus quality: 219831 bases at least Q20
Insert size: 220512; sum-of-contigs
Insert size: 221854; 2.3% error; agarose-fp
Quality coverage: 10.86x in Q20 bases; sum-of-contigs Quality
coverage: 11.12x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
source
1. 22212
   /organism="Mus musculus"
   /db_xref="taxon:10090"
   /chromosome="11"
   /clone="RP23-199H17"
   /clone_lib="RPC1-23"
1. 19515
   /note="assembly-fragment:03069
   /fragment_chain:1"
19616..27961
   /note="assembly-fragment:00400
   /fragment_chain:1"
28062..37082
   /note="assembly-fragment:03718
   /fragment_chain:1"
37183..45022
   /note="assembly-fragment:06200
   /fragment_chain:1"
45123..57322
misc_feature

```

	misc_feature	/note="assembly_fragment:00161 fragment_chain:1" 57423..59552 /note="assembly_fragment:02581 fragment_chain:1" 59653..77089 /note="assembly_fragment:00011 fragment_chain:1" 77190..88794 /note="assembly_fragment:05373 fragment_chain:1" 88895..101172 /note="assembly_fragment:05834 fragment_chain:1" 101273..112983 /note="assembly_fragment:02036 fragment_chain:1" 113084..119048 /note="assembly_fragment:00573 fragment_chain:1" 119149..139095 /note="assembly_fragment:04963 fragment_chain:1" 139196..141828 /note="assembly_fragment:02943 fragment_chain:1" 141929..150289 /note="assembly_fragment:04561 fragment_chain:2" 150390..166695 /note="assembly_fragment:06222 fragment_chain:2" 166796..169310 /note="assembly_fragment:04338 fragment_chain:2" 169411..210287 /note="assembly_fragment:05139 fragment_chain:3" 210388..222212 /note="assembly_fragment:02948 fragment_chain:3 clone_end:r7 vector_side:right"				
CASE COUNT	55706	a	53835	t	1706	others
ORIGIN			54341	g		

```

Query March      63.0%; Score 17; DB 2; Length 222212;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY      8  cytcaggtgcagptcag 24
      |||
Db 195872  CGTACGTGCAGGTTCAG 195888

```

RESULT	24
LOCUS	AE004050
DEFINITION	AE004050 10536 bp DNA BCT 15-JUN-2001
ACCESSION	AE004050 AE003849
VERSION	AE004050.1 GI:9107594
KEYWORDS	
SOURCE	<i>Xylella fastidiosa</i> 9a5c.
ORGANISM	<i>Xylella fastidiosa</i> 9a5c. Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group; <i>Xylella</i> .
REFERENCE	1 (bases 1 to 10536)
AUTHORS	Simpson,A.J., Reinach,F.C., Afrudga,P., Abreu,F.A., Acencio,M., Alvarenga,R., Alves,L.M., Araya,J.E., Bais,G.S., Baptista,C.S., Barros,M.H., Bonaccorsi,E.D., Bordi,S., Bove,J.M., Britones,M.R., Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carter,H., Colauto,N.B., Colomdo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,

TITLE	The genome sequence of the plant pathogen <i>Xylella fastidiosa</i> . The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing and Analysis
JOURNAL	Nature 406 (6792), 151-157 (2000)
MEDLINE	20365717
PUBMED	10910347
REFERENCE	2 (bases 1 to 10536)
AUTHORS	Simpson, A.J.G., Reischach, F.C., Arutud, P., Abreu, F.A., Acencio, M.,

Alaranga, R., Alves, L.M.C., Arya, J.E., Baia, G.S., Baptista, C.S., Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Britones, M.R.S., Bueno, M.R.P., Camargo, A.A., Camargo, L.E.A., Carraro, D.M., Carrier, H., Colauto, N.B., Colommo, C., Costa, F.F., Costa, M.C.R., Costa-Neto, C.M., Coutinho, L.L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorriy, H., Fachinchi, A.P., Ferreira, A.J.S., Ferreira, V.C.A., Ferro, J.A., Fraga, J.S., Franca, S.C., Franco, M.C., Fromme, M., Furlan, L.R., Garnier, M., Goldman, G.H., Goldman, H.S., Gomes, S.L., Gruber, A., Ho, P.L., Hohenseil, J.D., Junqueira, M.L.S., Kemper, E.L., Kitajima, J.P., Krieger, U.E., Kuramae, E.E., Laïret, F., Lamais, M.R., Leite, L.C.C., Lemos, E.G.M., Lemos, M.V.F., Lopes, S.A., Lopes, C.R., Machado, J.A., Machado, M.A., Madela, A.M.B.N., Madela, H.M.F., Marino, C.L., Marques, M.V., Martins, E.A.L., Martins, E.M.F., Matsukuma, A.Y., Menck, C.F.M., Miracca, E.C., Miyaki, C.Y., Monteiro-Vitorello, C.B., Moon, D.H., Nagai, M.A., Nascimento, A.L.T.O., Netto, L.E.S., Niani Jr., A., Nobrega, F.G., Nunes, L.R., Oliveira, M.A., de Oliveira, M.C., de Oliveira, R.C., Palmeri, D.A., Paris, A., Peixoto, B.R., Pereira, G.A.G., Pereira Jr., H.A., Pesquero, J.B., Quagno, R.B., Roberto, P.G., Rodrigues, V., de M. Rosa, A.J., de Rosa Jr., V.E., de Sa, R.G., Santelli, R.V., Sawasaki, H.E., da Silva, A.C.R., da Silva, F.R., da Silva, A.M., Silva Jr., W.A., da Silveira, J.F., Silvestri, M.L.Z., Siqueira, W.J., de Souza, A.A., de Souza, A.P., Tenenzi, M.F., Trufi, D., Tsai, S.M., Tsuchino, M.H., Vallada, H., Van Sluys, M.A., Verjovski-Almeida, S., Vettore, A.L.L., Zago, M.A., Zatz, M., Meidanis, J. and Setubal, J.C.

TITLE
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil

```

FEATURES
source
    1..10536
        location/Qualifiers
            /organism="Xylella fastidiosa 9a5c"
            /db_xref="taxon:160492"
            /clone="9a5c"
            1296..2339
                /gene="XF2409"
            1296..2339
                /gene="XF2409"
            /note="Similar to SPIP39369 (percent identity: 47 %/query
            alignment coverage: 96.8 %/subject alignment coverage:
            99.4 %); identified by sequence similarity; putative; ORF
            located using Glimmer/RBSfinder/Start codon shift: 2523"
            /codon_start=1
            /transl_table=11
            /product="hypothesized protein"
            /protein_id="AAF85208.1"
            /db_xref="GI:107595"
            /translation="MPALGVYHNDGLGKSGSGSRVLYEATIAAMLAENHESLQ
            KGERLHRTVGLTFPETOARATISQACRDVGLEVENHEDGIVGVHLDGKERVVLT
            SAVSKHADGGFTIDRASMLNVAVSRAKNFTLVFGDMVDFTPAKSRPRGLLAHRT
            DPSNALCFGLVRKDDLOQSTAVELVDAAEHDAFLDQALNVQREIHYSPTNPKRR
            IODIGAFAOMEAVEROYOVTAVTDDOLNTDDKKDKTKIAEVLQAAALRGVGEVHNP
            VDRVHSMKVIDDEDFVCVGSFNMSANRSMYAKHETSIVYRGKGLADEQRTLNLSLR
            ORITDDPQGVGQA"
            2311..2484
                /gene="XF2410"
            2311..2484
                /gene="XF2410"
            /note="hypothesized protein; identified by sequence

```



```

similarity: putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF85209.1"
/db_xref="GI:9107596"
/translation="MILSKSRADAVLRQRDVDTGWRQAMHAAEGRDWSPAVALRA
VALLKCCCLPLG"
complement(3181..5514)
/gene="XF2411"
complement(3181..5514)
/note="hypothetical protein: identified by sequence
similarity: putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF85210.1"
/db_xref="GI:9107597"
/translation="MLKKLIRVNHVGLIRNGVPTPTPPGPTVLFYAENGKCKTVA
NLRVAVHGKDKDEVLOKARTIDTDDPCIEILMNDGNTDITKYLHHGHWTGPPDL
IFDAVVDNVSQGEVHPAHKQFLQPVLGAEVSKLERETVLAKETADBNRIKFR
EORIRAYSKMNPIDAEIRLIPVDAEKIALRQREAVNAPALQORHAPQPLHIE
CNIDAFESIIOSTFHDIRADAEIIVQAHFOHQDPCIEAMVQGREFNQGCPCGC
QEVATALIKAYETVFNTRAFKEMKNVATGSGVATPEPEACELEKIDRIQDNACVA
AMKDOLDSPEFNPDPATATFTLRVALALASOKHRLDVCVGTODKADWVRLLE
ELRQMDAYNAAYVEINRITNFRTLAENTPALDLSATLEIGIYROSVEVQAT
VVOAKAKKEOLEEREKKNRMAALDAGLPDLSEYASKINQFLRDFGAFTSKRIQDSM
OGCTMRASVYLQLRGKEVALGRDSDGFHNVLSEDDKRTLALAFRLAYIPDL
VGSVYVLDPPMCSFDMTRNRMTESIALVMOGVQVVLSDHAYFLDLRLAGPY
MKVSNVHHIKRTKNDSOIVADADLITCSPYMLRACQAVTFSGTETGEOEVS
ALRPLVEGELKRFAPLLRQDLISIGOMISIRKATDPSPLVLAAPYVDLEKLNARA
VOSHHDKSLSIPMNDQOLROYARMALELIGGSLPH"
complement(5615..5800)
/gene="XF2412"
complement(5615..5800)
/note="hypothetical protein: identified by sequence
similarity: putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF85211.1"
/db_xref="GI:9107598"
/translation="MIFGATOTNADRRRSGTRHVACTGMAKSRLLPDPQVRYNAHLR
KPHYKVTSSIPITGML"
complement(6084..6512)
/gene="XF2413"
complement(6084..6512)
/note="hypothetical protein: identified by sequence
similarity: putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF85212.1"
/db_xref="GI:9107599"
/translation="MVGGDVPTADTRAAAGACSEPSIKNALRSCVHLLLCAGRLW
PGQCLGNGPMKGSVYKSSVYGLVLCSSCOQVEAAGAFARTTCKCMGMLWL
YRRWCSEHCLSCRTMEGATSGDAGFRGMLGKGRG"
complement(6509..8314)
/gene="XF2414"
complement(6509..8314)
/note="hypothetical protein: identified by sequence
similarity: putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF85213.1"
/db_xref="GI:9107600"
/translation="MPKRHOVHDAADDKENLYTTEHCFTGCVFTPLDSDGCFGRG
FSEPLAGLMLVMTIADVRYRRIQVAPLPLRMSPFVVAATGLTLTLLDMRAEQ

```

```

MLVPRGILTPAVMOALLGGAFLTLTWAMLGFTIRPPYGYQYNAARQOULYRILN
GSPTELAVIADLVSAEALVOHATNRKSLONGERTICEEEDDELPOVNDANALL
LIAKRLGAIIVAESPRTAFAIPALANKKGIPIRVAONIPEALANDEFLYHE
TNGCALGLIDGDKPFTQAMFAYEVETGTFENHDPRAKRWDPVOGLAYCSVYLT
LNSYAKQGRGTSEVLSQAMGTIASASDLSTLNEPSASDSVIRLALVOTFKDA
VRLDGEVBEHVCICVRAGYDSRAIFCDRIANNI FVYISLASEVRLPQLEMKIÖN
SWNSNLFHYGAFKRPAGQYVAFKLRRLRLNLAQDQRFENPERALIFECNVLAV
IRGNDYSDRRFHHRLHRAVLAKTRKHNVLHSHHPRILAESCLGNHTIYAEHLRAV
YHPAKGPICGKSVYFALNPASPPPEVVAASAVPNA"
complement(8365..8440)
/gene="trNA-Thr-2"
/note="found by tRNAscan"
/product="trNA-Thr"
complement(8365..8440)
/gene="trNA-Thr-2"
complement(8457..8648)
/gene="XF2415"
complement(8457..8648)
/note="hypothetical protein: identified by sequence
similarity: putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF85214.1"
/db_xref="GI:9107601"
/translation="MPVPRVAVDPPOPIILKSHLCATVAHLQSIQSPDPAIIMRRAP
EATFRKSSALSRRPSTAMM"
complement(8921..9871)
/gene="XF2416"
complement(8921..9871)
/note="similar to GI:5915671 (percent identity: 64 %/query
alignment coverage: 99.4 %/subject alignment coverage:
99.4 %); identified by sequence similarity: putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="drug tolerance protein"

```

Query Match 64.0%; Score 16; DB 1; Length 10536;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

8 cgtcagtcagtcagtcac 23
|||||
Db 1995 CGTCAGTCGACAGTCA 2010

RESULT 25
AC004508
LOCUS
DEFINITION Homo sapiens chromosome 21, p1 clone LBL#3 (LBNL H09), complete
SEQUENCE
AC004508 L35684 L35685 L35686 L35687 AC000939 AC000938 AC000940
AC000937 L81573 L35689 L35690 L35691 L35692 L35693 L35694 L35695
L35696 L35697 L35698 L35699 L42077 L35700 L35701
AC004508.1 GI:2996630
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K. S., Gunning, K. M.,
Davis, C. A., Kadner, K., Miguel, T., Pilchuck, S., Pollard, M.,
Rojeski, H., Subramanian, S., and Martin, C. H.
TITLE
JOURNAL
REFERENCE
AUTHORS
Ricke, D. O.
LARGE SCALE SEQUENCE ANALYSIS AND ANNOTATION WITH THE SEQUENCE
COMPARISON ANALYSIS (SCAN) SYSTEM


```

repeat_region complement(51869..52337)
                /rpt_family="MLT1"
repeat_region 52773..53082
                /rpt_family="Alu"
gene <53503..>67215
        /gene="MX2"
        /note="interferon-regulated resistance GTP-binding protein
        /evidence-experimental
        join(<53503..53668,54302..54494,56530..56664,58923..59077,
        67077..>67215)
CDS
        /gene="MX2"
        /note="interferon-regulated resistance GTP-binding protein
        /evidence-experimental
        /product="MX2"
        /protein_id="AAC08448.1"
        /db_xref="GI:2996631"
        /translation="FOOQPPCTVPPQMFPPNMOGAKEKDAFLAKDFNPLTLNNSQ
        /translating="FOOQPPCTVPPQMFPPNMOGAKEKDAFLAKDFNPLTLNNSQ
        PGNRSOPRAMEGNNLYSOYEQKRPCIDLDLSALGVEQDLALPALAVTODSSG
        KSVYALSGVAPRGSGIVTRCPVLYKLKKQCEAMAGRIYRNTELEQDQGVK
        EIRKXONVAGNGISHELISLEITSPEVDLTIIDLPITRVAADNDQPRDGLQIK
        ALIKKYTOROOTINLVVPCNVDIATTEALSMHEVDECDRTI"
repeat_region 57922..58215
                /rpt_family="Alu"
repeat_region 59599..59877
                /rpt_family="Alu"
repeat_region 60428..60810
                /rpt_family="TBE1"
repeat_region 61246..61366
                /rpt_family="Alu"

```

Query Match Best Local Similarity 100.0%; Score 16; DB 9; Length 71230;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 gtcaggtcaggtcag 24
DB 40455 gtcaggtcaggtcag 40470

RESULT 26
AC027810/c AC027810 155628 bp DNA HTG 26-MAY-2000
LOCUS Homo sapiens chromosome 18 clone RP11-357J12 map 18, WORKING DRAFT
DEFINITION
SEQUENCE, 14 unordered pieces.
AC027810
AC027810.3 GI:8082020
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 155628)
Homo sapiens chromosome 18, clone RP11-357J12
Unpublished
2 (bases 1 to 155628)
Birren,B., Linton,L., Nussbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArillano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferrelle,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hago,B., Heathford,A., Horton,L.,
Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Laroque,K., Lamazares,R., Landers,T., Lechoczky,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGuirk,A., McKernan,K., McPheters,R.,
Meldrum,J., Menus,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

TITLE JOURNAL COMMENT

O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessier,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Vieler,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:7684514.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: 357_J12
Center clone name: 19097

Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 146315 bases at least Q40
Consensus quality: 151241 bases at least Q40
Consensus quality: 153082 bases at least Q20
Insert size: 163000; agarose-fp
Insert size: 154328; sum-of-ctrls
Quality coverage: 4.2 in Q20 bases; agarose-fp
Quality coverage: 4.4 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1 1068: contig of 1068 bp in length
* 1069 1168: gap of 100 bp
* 1169 2717: contig of 1549 bp in length
* 2718 2817: gap of 100 bp
* 2818 6040: contig of 3223 bp in length
* 6041 6140: gap of 100 bp
* 6141 10201: contig of 4061 bp in length
* 10202 10301: gap of 100 bp
* 10302 14615: contig of 4314 bp in length
* 14616 14715: gap of 100 bp
* 14716 21086: contig of 6371 bp in length
* 21087 21186: gap of 100 bp
* 21187 25767: contig of 4581 bp in length
* 25768 25867: gap of 100 bp
* 25868 32500: contig of 6633 bp in length
* 32501 32600: gap of 100 bp
* 32601 38651: contig of 6051 bp in length
* 38652 38751: gap of 100 bp
* 38752 50922: contig of 12171 bp in length
* 50923 51022: gap of 100 bp
* 51023 71014: contig of 19992 bp in length
* 71015 71114: gap of 100 bp
* 71115 93563: contig of 22449 bp in length
* 93564 93663: gap of 100 bp
* 93664 123080: contig of 29417 bp in length
* 123081 123180: gap of 100 bp
* 123181 155628: contig of 32448 bp in length.
Location/Qualifiers
1..155628
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

FEATURES SOURCE

```

/misc_feature /chromosome="18"
/misc_feature /map="18"
/misc_feature /clone="RP11-357J12"
/misc_feature /clone_lib="RP11-11 Human Male BAC"
/misc_feature 1..1068
/misc_feature /note="assembly_fragment"
/misc_feature 1169..2217
/misc_feature /note="assembly_fragment"
/misc_feature 2818..6040
/misc_feature /note="assembly_fragment"
/misc_feature 6141..10201
/misc_feature /note="assembly_fragment"
/misc_feature 10302..14615
/misc_feature /note="assembly_fragment"
/misc_feature 14716..21086
/misc_feature /note="assembly_fragment"
/misc_feature 21187..25767
/misc_feature /note="assembly_fragment"
/misc_feature 25868..32500
/misc_feature /note="assembly_fragment"
/misc_feature 32601..38651
/misc_feature /note="assembly_fragment"
/misc_feature 38752..50922
/misc_feature /note="assembly_fragment"
/misc_feature 51023..71014
/misc_feature /note="assembly_fragment"
/misc_feature 71115..93563
/misc_feature /note="assembly_fragment"
/misc_feature 93664..123080
/misc_feature /note="assembly_fragment"
/misc_feature 123181..155628
/misc_feature /note="assembly_fragment"
misc_feature clone_end:SP6
misc_feature vector:side:left"

BASE COUNT 37665 a 38986 c 40608 g 37067 t 1302 others
ORIGIN

Query Match 64.0% Score 16; DB 2; Length 155628;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 gtcaggtcaggtcag 24
|||||
Db 95800 GTCAGTCAAGTCAG 95785

RESULT 27
AL139326/c 175588 bp DNA PRI 09-MAY-2001
LOCUS Human DNA sequence from clone RP11-351K3 on chromosome 13, complete
DEFINITION
ACCESSION AL139326
VERSION AL139326.15 GI:14018249
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 175588)
AUTHORS Blakey, S.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT
On May 11, 2001 this sequence version replaced gi:13277016.
During sequence assembly data is compared from overlapping clones.
These differences are found there are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all

```

regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL, SW, SWISSPROT, Tr, TrEMBL, Wp, WormRep; Information on the WormRep database can be found at <http://www.sanger.ac.uk/projects/C-elegans/wormrep> This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr13> RP11-351K3 is from the library RP11-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-351K3. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP11-351K3 is at 1 in this sequence. The true left end of clone RP11-154C3 is at 175489 in this sequence.

FEATURES

source

repeat_region 1..175588
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-351K3"
/clone_lib="RP11-11.2"
68..360
/note="Alusx repeat: matches 1..290 of consensus"
953..1089
/note="MIR repeat: matches 80..210 of consensus"
3306..3437
/note="L1MD1 repeat: matches 6066..6201 of consensus"
3591..3719
/note="MER47 repeat: matches 1..193 of consensus"
3722..3797
/note="MER47 repeat: matches 2248..2323 of consensus"
3803..3878
/note="L1MFC repeat: matches 2203..2279 of consensus"
3879..4182
/note="Alu repeat: matches 1..289 of consensus"
4183..4479
/note="L1MFC repeat: matches 1903..2203 of consensus"
4480..4788
/note="Alusg repeat: matches 2..310 of consensus"
4789..5617
/note="L1MFC repeat: matches 1094..1903 of consensus"
5628..5925
/note="Alu repeat: matches 1..298 of consensus"
6403..6459
/note="MIR repeat: matches 197..259 of consensus"
6899..7218
/note="Alu repeat: matches 1..311 of consensus"
7225..7364
/note="MIR repeat: matches 46..188 of consensus"
7961..8260
/note="Alub repeat: matches 1..300 of consensus"
8440..8736
/note="Alub repeat: matches 1..298 of consensus"
9747..9956
/note="MIR repeat: matches 36..250 of consensus"
10124..10315
/note="L1ME1 repeat: matches 5988..6322 of consensus"
10384..10682
/note="Alu repeat: matches 1..300 of consensus"
10954..11235
/note="Alu repeat: matches 1..294 of consensus"
11236..11543
/note="Alus repeat: matches 3..294 of consensus"

```

repeat_region 12286..12493
/note="MIR repeat: matches 32..249 of consensus"
repeat_region 13328..13622
/note="Alusx repeat: matches 1..308 of consensus"
repeat_region 15766..15864
/note="L2 repeat: matches 2630..2736 of consensus"
repeat_region 16580..16778
/note="L2 repeat: matches 2540..2742 of consensus"
repeat_region 17117..17212
/note="MIR repeat: matches 30..124 of consensus"
repeat_region 17358..18333
/note="SV4 repeat: matches 2..954 of consensus"
misc_feature 17886..19097
/note="CpG Island"
evidence-not-experimental
18166..18189
/note="1372 bases of 1S186 (X03123) removed here. This
sequence represents the duplicated flanking sequence of
the 1S186."
repeat_region 18343..19109
/note="SV4 repeat: matches 646..1386 of consensus"
repeat_region 19813..21277
/note="L1M4 repeat: matches 4785..6297 of consensus"
repeat_region 21663..21969
/note="Alub repeat: matches 1..308 of consensus"
repeat_region 22425..22651
/note="L1M4 repeat: matches 4456..4697 of consensus"
repeat_region 23061..23181
/note="MIR repeat: matches 34..170 of consensus"
repeat_region 23287..23332
/note="L2 repeat: matches 2692..2737 of consensus"
repeat_region 23293..23345
/note="MIR repeat: matches 200..249 of consensus"
repeat_region 26800..27101
/note="Alusg repeat: matches 1..302 of consensus"
repeat_region 27275..27582
/note="Alub repeat: matches 1..306 of consensus"
repeat_region 27607..27900
/note="AlusC repeat: matches 1..308 of consensus"
repeat_region 27909..28044
/note="FLAMC repeat: matches 1..133 of consensus"
repeat_region 28130..28431
/note="Alub repeat: matches 6..305 of consensus"
repeat_region 28517..28572
/note="MIR repeat: matches 96..155 of consensus"
repeat_region 29159..29275
/note="Alub repeat: matches 1..118 of consensus"
repeat_region 29425..29729
/note="Alu repeat: matches 1..305 of consensus"
repeat_region 30098..30173
/note="L1M5 repeat: matches 7840..7913 of consensus"
repeat_region 30259..30544
/note="Alub repeat: matches 1..284 of consensus"
repeat_region 30739..30875
/note="L1M5 repeat: matches 7603..7731 of consensus"
repeat_region 30860..31103
/note="L1M4 repeat: matches 7643..7901 of consensus"
repeat_region 31112..31223
/note="Alusg repeat: matches 194..305 of consensus"
repeat_region 31252..31598
/note="L1M4 repeat: matches 4649..4974 of consensus"
repeat_region 31646..32054
/note="L1M4 repeat: matches 6419..6828 of consensus"
repeat_region 32497..32783
/note="Alus repeat: matches 1..291 of consensus"
repeat_region 32784..32851
/note="34 copies 2 mer tg 8% conserved"
repeat_region 32852..32873
/note="11 copies 2 mer ta 100% conserved"
repeat_region 33312..34465
/note="L1M3 repeat: matches 6641..7739 of consensus"
repeat_region 34501..34531
/note="L1M3 repeat: matches 5494..5524 of consensus"

```

```

repeat_region 35395..35581
/note="L1P2 repeat: matches 5958..6152 of consensus"
repeat_region 37300..37451
/note="L2 repeat: matches 1929..2081 of consensus"
repeat_region 37499..37661
/note="MER33 repeat: matches 8..156 of consensus"
repeat_region 37662..37968
/note="Alusx repeat: matches 1..307 of consensus"
repeat_region 37969..38061
/note="MER33 repeat: matches 156..243 of consensus"
repeat_region 38063..38361
/note="Alusx repeat: matches 1..296 of consensus"
repeat_region 38434..38651
/note="Alu repeat: matches 1..217 of consensus"
repeat_region 38654..38833
/note="Alusg/x repeat: matches 121..299 of consensus"
repeat_region 38947..39134
/note="L2 repeat: matches 1470..1655 of consensus"
repeat_region 40240..40544
/note="Alusx repeat: matches 1..305 of consensus"
repeat_region 40547..40855
/note="Alusg repeat: matches 3..310 of consensus"
repeat_region 41027..41160
/note="L1M4 repeat: matches 7837..7977 of consensus"
repeat_region 41161..41198
/note="19 copies 2 mer ac 86% conserved"
repeat_region 41204..41241
/note="19 copies 2 mer tg 81% conserved"
repeat_region 41252..41356
/note="L1M4 repeat: matches 7718..7832 of consensus"

```

```

Query Match 64.0%: Score 16; DB 9; Length 175588;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 10 tcagtgacagtcacg 25
Db 9769 TCAGTGCAGTCAGC 9754

```

```

RESULT 28
AC005412/c DNA PRI 22-NOV-2000
LOCUS Homo sapiens chromosome 17, clone hRPK.22_N.12, complete sequence.
AC005412
ACCESSION AC005412
VERSION AC005412.6 GI:11128436
KEYWORDS HPG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euteria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 183700)
REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone hRPK.22_N.12
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 183700)
Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J.,
Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,M.,
Collins,S., Collumore,A., Cooke,P., Corliss,D., Depayre,E.,
Devon,K., Dewar,K., Donelan,L., Ferreira,P., Fitzhugh,W.,
Forester,C., Funke,R., Gage,D., Gardina,S., Geradgery,K., Grant,G.,
Hagos,B., Hearford,A., Herena,L., Horton,L., Howland,J.C.,
Jacrot,L., Jones,C., Kann,L., Karatas,A., Lehotzky,J.,
Macdonald,P., Margulis,N., McEwan,P., McGurk,A., McKernan,K.,
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,D.,
Nahf,R., Naylor,J., Nilot,M., O'Connor,T., O'Donnell,P.,
Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
Strange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,
Subramanian,A., Tesfaye,S., Tichonovskiy,N., Tortuella-Miller,I.,
Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D.,
Ye,W.-J., Zhao,J. and Zody,M.
Direct Submission

```

TITLE

JOURNAL	Submitted (12-AUG-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	repeat_region	/rpt_family="AluJb"
REFERENCE	3 (bases 1 to 183700)	repeat_region	complement(4248..4361)
AUTHORS	Birden,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldini,J., Barna,N., Beckerly,R., Boguslavsky,L., Bouknight,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczy,J., Liu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McDonald,J., Meldrum,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Tittell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.D., Zimmer,A. and Zody,M.	repeat_region	/rpt_family="AluJb"
TITLE	Direct Submission	repeat_region	complement(4248..4361)
JOURNAL	Submitted (29-SEP-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	repeat_region	/rpt_family="AluJb"
REFERENCE	4 (bases 1 to 183700)	repeat_region	complement(4362..4490)
AUTHORS	Birden,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldini,J., Barna,N., Beckerly,R., Boguslavsky,L., Bouknight,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczy,J., Liu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McDonald,J., Meldrum,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Tittell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.D., Zimmer,A. and Zody,M.	repeat_region	complement(4493..4589)
TITLE	Direct Submission	repeat_region	complement(4590..4888)
JOURNAL	Submitted (09-NOV-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	repeat_region	complement(4889..5305)
COMMENT	On Nov 9, 2000 this sequence version replaced g1:5931398. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html .	repeat_region	complement(5328..5623)
FEATURES	Location/Qualifiers	repeat_region	complement(5676..6164)
source	1..183700	repeat_region	complement(6164..6509)
	/organism="Homo sapiens"	repeat_region	/rpt_family="AluSg"
	/db_xref="taxon:9606"	repeat_region	complement(6580..6784)
	/map="17"	repeat_region	/rpt_family="L2"
	/clone_lib="RPCT-11 Human Male BAC"	repeat_region	complement(6986..7137)
	1..70	repeat_region	complement(7138..7447)
	/rpt_family="MER41D"	repeat_region	complement(7448..7562)
	784..1084	repeat_region	complement(7562..7808)
	/rpt_family="AluY"	repeat_region	/rpt_family="AluSg"
	1320..1637	repeat_region	/rpt_family="AluJb"
	/rpt_family="AluJc"	repeat_region	complement(7825..8084)
	complement(1664..1959)	repeat_region	complement(8084..8121)
	1961..2074	repeat_region	complement(8121..8476)
	/rpt_family="L2"	repeat_region	complement(8476..8677)
	2283..2355	repeat_region	complement(8677..8931)
	/rpt_family="C-rich"	repeat_region	complement(8931..9210)
	complement(2369..2774)	repeat_region	complement(9210..9395)
	/rpt_family="L1M4"	repeat_region	complement(9395..9530)
	2775..3066	repeat_region	complement(9530..9878)
	/rpt_family="AluSg"	repeat_region	complement(9878..10115)
	complement(3067..3077)	repeat_region	complement(10115..10128)
	/rpt_family="L1M4"	repeat_region	complement(10128..10357)
	3549..3814	repeat_region	complement(10357..10432)
	/rpt_family="AluSg"	repeat_region	complement(10432..10458)
	complement(3867..3949)	repeat_region	complement(10458..10511)
	/rpt_family="AluJc"	repeat_region	complement(10511..10567)
	complement(3950..4247)	repeat_region	complement(10567..10908)
		repeat_region	complement(10908..10913)
		repeat_region	complement(10913..11210)
		repeat_region	complement(11210..11512)
		repeat_region	complement(11512..11529)
		repeat_region	complement(11529..11759)
		repeat_region	complement(11759..12159)
		repeat_region	complement(12159..12159)

```

repeat_region complement(13008..13306)
repeat_region /rpt_family="AlusP"
repeat_region 13355..13490
repeat_region /rpt_family="AlusX"
repeat_region 13491..13789
repeat_region /rpt_family="AlusC"
repeat_region 13790..13956
repeat_region /rpt_family="AlusX"
repeat_region complement(14078..14206)
repeat_region /rpt_family="MIR"
repeat_region complement(14220..14276)
repeat_region /rpt_family="L2"
repeat_region complement(14277..14309)
repeat_region /rpt_family="MIR"
repeat_region complement(14310..14594)
repeat_region /rpt_family="AlusP"
repeat_region complement(14604..14912)
repeat_region /rpt_family="AlusO"
repeat_region complement(14913..15045)
repeat_region /rpt_family="MIR"
repeat_region complement(15061..15157)
repeat_region /rpt_family="MIR"
repeat_region complement(15517..15637)

```

```

Query Match 64.0% Score 16: DB 9: Length 183700:
Best Local Similarity 100.0%: Pred. No. 35:
Matches 16: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

```

```

Oy 9 gtcaggtacagtcag 24
11111111111111111111
Db 177274 GTCAGGTACAGTCAG 177259

```

```

RESULT 29
LOCUS AE001862/c DNA 22-NOV-1999
DEFINITION Deinococcus radiodurans RI section 1 of 2 of the complete
chromosome 2
ACCESSION AE001862 AE001825
VERSION AE001862.1 GI:6460468
KEYWORDS
SOURCE
ORGANISM Deinococcus radiodurans.
Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.

```

```

REFERENCE 1 (bases 1 to 213732)
AUTHORS White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,
Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zaleski, C.,
Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M., et al.
Genome sequence of the radioresistant bacterium Deinococcus
radiodurans RI
JOURNAL Science 286 (5444), 1571-1577 (1999)
MEDLINE 2 (bases 1 to 213732)
20036896
AUTHORS White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,
Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zaleski, C.,
Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W.,
Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S.,
Smith, H.O., Venter, J.C. and Fraser, C.M.
Direct submission

```

```

TITLE Submitted (02-NOV-1999) The Institute for Genomic Research, 9712
JOURNAL Medical Center Dr, Rockville, MD 20850, USA

```

```

FEATURES
source
1. 213732
/organism="Deinococcus radiodurans"
/strain="RI"
/db_xref="taxon:1299"
/chromosome="2"
653..1435
/gene="DRA0001"

```

CDS

```

653..1435
/gene="DRA0001"
/feature="similar to GB:D26185.SP:P37522 PID:467381
PID:580906 GB:AL009126 percent identity: 54.96; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="chromosome partitioning ATPase, putative, para
family"
/protein_id="AAFI2301.1"
/db_xref="GI:6460595"
/transl_table=11
/translation="MVSAYKTLTVFNHAGAGKTSLTNLNGVELARGLRVLIDLP
QANLTGMLGISGVTREKTVYPAVDGGLPSPVAFAGLDVTPAHVSLAVAGOMGKRY
CAOGRRLRALAEVSGDVALDPSLGAOLAIALAAQOMIVPTROKGLDLP
LOGALTEYREVPDLTVLALVPTPEARRRHDOELADLKAHLSPLARVPQGRVAVML
DSTAGAPVSETPPVHADYORLTADIAAIGVAYGENA"
1432..2313
/gene="DRA0002"
1432..2313
/gene="DRA0002"
/feature="similar to GB:134077 GB:X85964 PID:407375
PID:757761 PID:1163135 percent identity: 57.34; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="chromosome partitioning protein, parB family"
/protein_id="AAFI2300.1"
/db_xref="GI:6460594"
/transl_table=11
/translation="MTRRRPERRDLLGLGETPYDLSQANDIRALPYNEIKVSTOP
RRSFDERLSLAEISIRAGVLOPLVRSVDQYEIVAGERRRAQLAGLAEVAVV
RQLSNEQAAALINLQDNLVNIDVGGKLELALTLGLREBARRLQHLRAVP
GDEHOLDQVFRSMGFTWTFPAKNTKLINMPOPLVLEALRGLPTLSVVAAPER
QNELKIANGASRSQILQALOTPSQTSVAVPEHAKYLSKRFLSGLDPTREALDR
WLAKEPERVROAIDEGS"
complement(2418..3881)
/gene="DRA0003"
complement(2418..3881)
/feature="similar to GB:M88334.SP:P25526 PID:147901 GB:U00096
PID:1789015 percent identity: 69.60; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="succinate-semialdehyde dehydrogenase"
/protein_id="AAFI2294.1"
/db_xref="GI:6460588"
/transl_table=11
/translation="MTTDLRTTYSVTRSCAYPDGEMRNAPRNFVHPNGEYIGE
VADCPPTDARQIDAEEVALREROVNPEYERKILRWMDLMEEREEIADLMTLMEG
KPISETRGEVHYAASFIEWCAEEAGRIAGERINLREPHKRGILTISEPGIVYAVTPMN
PAGMTITRAAPALAGCYMILKPAELISPTALYLTLEMLKAGNAPFVQVNDAS
ALTQPPMNSRYRKLTFTGSTEVGRLLYQQAAGTIRKYSLELGGHAPFVDDADLER
AASEYVASKFRNSGCCTNRYVORGVAAEFIRLLTAAQLDGPDPDAPVCP
VVEOAGLDKVVOROVADALTKGAATTTGGVSSGLFPQPLVDVAPDSILRETFGP
VAEVTITDTEEGRLANDSEGLAAVATYTRDGRFRAEGLEIVGINDLPSSA
APHPFGMKNSGVREGHGMGLEYLETKFVSLGS"
4024..4668
/gene="DRA0004"
4024..4668
/gene="DRA0004"
/feature="similar to GP:2924378 percent identity: 55.29;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAFI2299.1"
/db_xref="GI:6460593"
/transl_table=11
/translation="MNSLSLAIVLLMAGSTRMGOQKALPLGMPICRHAQTVLE
AGGEVWAVLPBETGDAVRAALQGLPLWVTPNPQAGLSSFOAGGALVAESAAP
AAVFTLADMLVARDTHQALGVGAEGPAAITIRGEVAAAPTLRLRLPLRLEFP
PSDHGRLLIRELGRKLIKIDRPAALDIDRPEMAAQOQLFAVTEGAQGGP"
complement(4719..5897)
/gene="DRA0005"
complement(4719..5897)

```

CDS

```

653..1435
/gene="DRA0001"
/feature="similar to GB:D26185.SP:P37522 PID:467381
PID:580906 GB:AL009126 percent identity: 54.96; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="chromosome partitioning ATPase, putative, para
family"
/protein_id="AAFI2301.1"
/db_xref="GI:6460595"
/transl_table=11
/translation="MVSAYKTLTVFNHAGAGKTSLTNLNGVELARGLRVLIDLP
QANLTGMLGISGVTREKTVYPAVDGGLPSPVAFAGLDVTPAHVSLAVAGOMGKRY
CAOGRRLRALAEVSGDVALDPSLGAOLAIALAAQOMIVPTROKGLDLP
LOGALTEYREVPDLTVLALVPTPEARRRHDOELADLKAHLSPLARVPQGRVAVML
DSTAGAPVSETPPVHADYORLTADIAAIGVAYGENA"
1432..2313
/gene="DRA0002"
1432..2313
/gene="DRA0002"
/feature="similar to GB:134077 GB:X85964 PID:407375
PID:757761 PID:1163135 percent identity: 57.34; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="chromosome partitioning protein, parB family"
/protein_id="AAFI2300.1"
/db_xref="GI:6460594"
/transl_table=11
/translation="MTRRRPERRDLLGLGETPYDLSQANDIRALPYNEIKVSTOP
RRSFDERLSLAEISIRAGVLOPLVRSVDQYEIVAGERRRAQLAGLAEVAVV
RQLSNEQAAALINLQDNLVNIDVGGKLELALTLGLREBARRLQHLRAVP
GDEHOLDQVFRSMGFTWTFPAKNTKLINMPOPLVLEALRGLPTLSVVAAPER
QNELKIANGASRSQILQALOTPSQTSVAVPEHAKYLSKRFLSGLDPTREALDR
WLAKEPERVROAIDEGS"
complement(2418..3881)
/gene="DRA0003"
complement(2418..3881)
/feature="similar to GB:M88334.SP:P25526 PID:147901 GB:U00096
PID:1789015 percent identity: 69.60; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="succinate-semialdehyde dehydrogenase"
/protein_id="AAFI2294.1"
/db_xref="GI:6460588"
/transl_table=11
/translation="MTTDLRTTYSVTRSCAYPDGEMRNAPRNFVHPNGEYIGE
VADCPPTDARQIDAEEVALREROVNPEYERKILRWMDLMEEREEIADLMTLMEG
KPISETRGEVHYAASFIEWCAEEAGRIAGERINLREPHKRGILTISEPGIVYAVTPMN
PAGMTITRAAPALAGCYMILKPAELISPTALYLTLEMLKAGNAPFVQVNDAS
ALTQPPMNSRYRKLTFTGSTEVGRLLYQQAAGTIRKYSLELGGHAPFVDDADLER
AASEYVASKFRNSGCCTNRYVORGVAAEFIRLLTAAQLDGPDPDAPVCP
VVEOAGLDKVVOROVADALTKGAATTTGGVSSGLFPQPLVDVAPDSILRETFGP
VAEVTITDTEEGRLANDSEGLAAVATYTRDGRFRAEGLEIVGINDLPSSA
APHPFGMKNSGVREGHGMGLEYLETKFVSLGS"
4024..4668
/gene="DRA0004"
4024..4668
/gene="DRA0004"
/feature="similar to GP:2924378 percent identity: 55.29;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAFI2299.1"
/db_xref="GI:6460593"
/transl_table=11
/translation="MNSLSLAIVLLMAGSTRMGOQKALPLGMPICRHAQTVLE
AGGEVWAVLPBETGDAVRAALQGLPLWVTPNPQAGLSSFOAGGALVAESAAP
AAVFTLADMLVARDTHQALGVGAEGPAAITIRGEVAAAPTLRLRLPLRLEFP
PSDHGRLLIRELGRKLIKIDRPAALDIDRPEMAAQOQLFAVTEGAQGGP"
complement(4719..5897)
/gene="DRA0005"
complement(4719..5897)

```

```

/gene="DRA0005"
/Note="similar to GB:L33464 SP:P47734 PID:496118 percent
identity: 74.23; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="alcohol dehydrogenase, zinc-containing"
/protein_id="AAFI2293.1"
/db_xref="GI:6460587"
/translation="MKAIWGVNKKVVEVTPDPTLLPDAIVKITSIAICGSDHL
LDGYTPMETDILGHEFMGVGVSEVVKLKAGDVRVVPFNACGVCPPCKRFEFS
ACDQNSNHNHMAEALYATSGALFSGSHMGAGVAGGQAVRVPADVPEFIESL
KDEQVLELIDFPTGYOAECCNIVPGRDVAVGAGPVGAPARSOMTGAHIVYI
DRVEERLMAEACGCTINYEKEDVLLSTREATGGRGPDHIVDAVMEAGHCGSTM
DKVOONLKTEDRTALRMATLSACKCTYSMEVYGLVDKMPIGAAFAKGLIFRMG
QTHRHHPILSLGRIEAGQIDPFEVITTHRASLDAPELVATFERKQDCIKVLPNPA
"
complement(5894..6700)
/gene="DRA0006"
complement(5894..6700)
/gene="DRA0006"
/Note="similar to PID:1653488 percent identity: 48.46;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAFI2292.1"
/db_xref="GI:6460586"
/translation="MDRRSSPOLRTMTNMNREOPADPHACATCNSSPCATMPLERSI
TGAVGALAAAGLRKSRGLERVVLTGAGALIVAAAGQNPATALKTEQNDGEIVV
RDVAITGRPAEQLYAVWRDPLGLMTHTLSVEVLEDDKSRMTVEAPAVLGSVEA
ELTADPEGRITAMRSLEPGARIEGSEVLPFPARGARTEVVVRLTYRPPGSGAVIA
RMFQEPSCQRLRDLRMFRKKEQLGHAPTAQCGSSGSAALTPPNKGDLY"
6839..7744
/gene="DRA0007"
6839..7744
/gene="DRA0007"
/Note="similar to GB:AL009126 percent identity: 62.98;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ABC transporter, ATP-binding protein, putative"
/protein_id="AAFI2291.1"
/db_xref="GI:6460585"
/translation="MITTEOVSKTYGHVATLSDFNLRTGELTALLCPNGAKSTIA
IGLLIGSASACQVRLGADPRNDVRARIGAPQESALPGLTREVAVTI.FASYP
APLVDEALALADLGPVAGRAOALSGGGRRLAFALVAGDEPILLIDEPYGMQO
SRAFWAVATVGLARGRTILLTHYIEAEERTADRVVNMGGIITADIPQGRSGVC
GARVSPSDLYQAELERLPGVSAVOYDAGRADLRVSPALLALILGSGTTFSDLEV
KRAILEVATVQLTGPDMTAVTRSA"
7861..8652
/gene="DRA0008"
7861..8652
/gene="DRA0008"
/Note="similar to GB:AL009126 percent identity: 52.46;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAFI2290.1"
/db_xref="GI:6460584"
/translation="MTTLNVPARRTSALPVLVLAQLTGLERKLNSPMTVGATGFPV
MFPAHGLPAVEYCATDPHVGVTITQFNAVSLSLAFSGAVALTERSGMRLI.
RSSPLPVPPLYPVASKTALAFAVSLALYFAHFAHFGVTLPLGLALLAGLILGMI
PLVALGICIGFLASPOAOILANILSVMSFASCPALPDLPGFQOQIAPLLPAYHV
SCIAITVSGGTASAPAHWALAAFTLVGTILVWMLKRDSERES"
8809..9819
/gene="DRA0009"
8809..9819
/gene="DRA0009"
/Note="similar to GB:AL009126 percent identity: 55.80;
identified by sequence similarity; putative"
/codon_start=1

```

```

/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAFI2289.1"
/db_xref="GI:6460583"
/translation="MLFNGTIVAPAVVFFFGKRPDDDDRAVYGVWSALIMMAGDM
LTGPSALSLVYTGSLIGWQRQWALCGAAILNALMVGRIITLPERSQGFIFLFT
LASAVANHAITYGLARRRLAVALQVLEKLAADARERLADLDLHGLHTSVILKS
ELAGKLAEKHPORAAAEI REVERISREALSSEVRAVQGRSGSLAAELARAVLADA
GIRLVITDPLPELPADLESTLAMVIREAVTVNVRHSGASELRLTIRGANGFOLTIOD
DGRGADPAGTGTLNSMRRLRAMGTFFPRDGTGRIQAFVPLASGTAGLKAPELLEK
"
9816..10451
/gene="DRA0010"
9816..10451
/gene="DRA0010"
/Note="similar to GB:AL009126 percent identity: 67.84;
Query Match 64.0%; Score 16; DB 1; Length 213732;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 tcaggtcaggtcagc 25
Db 52231 TCAGGTGACAGTCAGC 52216
RESULT 30
SME591789/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Sinorhizobium meliloti.
Sinorhizobium meliloti
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Sinorhizobium.
1 (bases 1 to 294800)
Capela,D., Barloy-Hubler,F., Gouzy,J., Botte,G., Ampe,F., Batut,J.,
Boistard,F., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
Godt,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,
Pohl,T., Portetlelle,D., Puehler,A., Purnelle,B., Ramsperger,U.,
Renard,C., Thebaud,P., Vandenbol,M., Weidner,S. and Galibert,F.
From the Cover: Analysis of the chromosome sequence of the legume
symbiont Sinorhizobium meliloti strain 1021
Proceedings of the National Academy of Sciences of the United
States of America. 98 (17), 9877-9882 (2001)
11481430
2 (bases 1 to 294800)
Gouzy,J.
Direct Submission
Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO
EU Consortium
MELILO EU Consortium:
Laboratoire de Biologie Molculaire des Relations
Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
France, Laboratoire de Genetique et Developpement UMR6061-CNRS,
Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,
France, GATC GmbH, Fritz-Arnold-Str. 23, D-78467 Konstanz, Germany,
Universitaet Bielefeld, Biologie IV (Genetik) Universitaetsstr 25,
D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,
Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20,
B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte
des Sciences Agronomiques de Gembloux, Avenue Mareschal Juin 6,
B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@loulouise.inra.fr
http://sequence.loulouise.inra.fr/meliloti.html.
Location/Qualifiers
1..294800
/organism="Sinorhizobium meliloti"
/strain="1021"
/db_xref="taxon:382"
complement(1147..11382)
/gene="Skc04259"
FEATURES
source
gene

```


CDS
complement(147..1382)
/gene="SMC04259"
/function="cell processes: transport of small molecules"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE PERIPLASMIC BINDING ABC TRANSPORTER
SIGNAL PEPTIDE PROTEIN"
/protein_id="CAC46508.1"
/db_xref="GI:15074951"
/translation="MNLRLAALGATLPGASATDLEVTIMHTSGGEAAVAEL
AKAFDITGNKWDGAIAGSGTARPTIMISRTIGDPMATQENHQAELVQAGMR
DLTDITATKEMWKEIVKPSLLDSCITEGKLYCAPVTHSMQMLSNAAFKAQGEVP
KNMDEFVAAPALEKAGIYPLVAGGQPMQANGADVLVYAIAGKENEFVAFQDEEV
AAGPEIAKVKRKADARBMKSGTNVDMMQATNMVITGKAGGQIMGDMQOGEFOLAGO
KACVDYTCPLGICVNEVISTGDAFYFPLLEDEKSKAQEVLASTLLKPEYQAFMLK
KGLPLVGDVDLAANDCKKGLDILAKGNVIOGTQDILLSADSKOKEDLSEFFANH
SMTPEAQKRFADILAAAD"
1858..2928
/gene="SMC04260"
1858..2928
/gene="SMC04260"
/function="miscellaneous; not classified regulator"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAC46509.1"
/db_xref="GI:15074952"
/translation="MTGMDGKTKNRATTAPEVGGRRPLTKTIAFMGLIGVTVSPRAK
DADIGAEIKERYRLVAKQIGOPNRAGVLRFGKRTNVISLVITLPEETITSPKVI
GITEILAGTGYHLVTPYSSAKDPLGPIRILDTGADGVYISRTPEENDRVLLTER
RLPEATGRTGEMGLIHPYHDNREFAVEARKLVDRGRRLVLEPPMLTJHSHNR
TGFEGVGRDGAELVSPHQVNDISLVAIRDAEKEIMRSPDADGVVSGSGSAIALI
AGTEAACKKVGDEAEMVSKVPSDFRLMLREVTMTYEDIHAGRELAKAVIGHGEP
PERTQSLSEPEFOPMARSAVS"
2954..3023
/note="Sm-4 OR SMC04506
REPEAT SM-4
predicted by Homology"
/evidence=not_experimental
complement(3036..3302)
/gene="SMC04261"
complement(3036..3302)
/gene="SMC04261"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="HYPOTHEICAL PROTEIN"
/protein_id="CAC46510.1"
/db_xref="GI:15074953"
/translation="MKFTVLEGKAAEVTAAGSVTIDAEERRIKAAQYDQWRVRSIA
TGAPIPGIQYLKMQIDVAVAGKLEOLNIPSDFTDKYWPAGI"
gene
complement(3407..4837)
/gene="gnd OR SMC04262"
complement(3407..4837)
/gene="gnd OR SMC04262"
/EC_number="1.1.1.44"
repeat_region
gene
complement(147..1382)
/gene="SMC04259"
/function="cell processes: transport of small molecules"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE PERIPLASMIC BINDING ABC TRANSPORTER
SIGNAL PEPTIDE PROTEIN"
/protein_id="CAC46508.1"
/db_xref="GI:15074951"
/translation="MNLRLAALGATLPGASATDLEVTIMHTSGGEAAVAEL
AKAFDITGNKWDGAIAGSGTARPTIMISRTIGDPMATQENHQAELVQAGMR
DLTDITATKEMWKEIVKPSLLDSCITEGKLYCAPVTHSMQMLSNAAFKAQGEVP
KNMDEFVAAPALEKAGIYPLVAGGQPMQANGADVLVYAIAGKENEFVAFQDEEV
AAGPEIAKVKRKADARBMKSGTNVDMMQATNMVITGKAGGQIMGDMQOGEFOLAGO
KACVDYTCPLGICVNEVISTGDAFYFPLLEDEKSKAQEVLASTLLKPEYQAFMLK
KGLPLVGDVDLAANDCKKGLDILAKGNVIOGTQDILLSADSKOKEDLSEFFANH
SMTPEAQKRFADILAAAD"
1858..2928
/gene="SMC04260"
1858..2928
/gene="SMC04260"
/function="miscellaneous; not classified regulator"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAC46509.1"
/db_xref="GI:15074952"
/translation="MTGMDGKTKNRATTAPEVGGRRPLTKTIAFMGLIGVTVSPRAK
DADIGAEIKERYRLVAKQIGOPNRAGVLRFGKRTNVISLVITLPEETITSPKVI
GITEILAGTGYHLVTPYSSAKDPLGPIRILDTGADGVYISRTPEENDRVLLTER
RLPEATGRTGEMGLIHPYHDNREFAVEARKLVDRGRRLVLEPPMLTJHSHNR
TGFEGVGRDGAELVSPHQVNDISLVAIRDAEKEIMRSPDADGVVSGSGSAIALI
AGTEAACKKVGDEAEMVSKVPSDFRLMLREVTMTYEDIHAGRELAKAVIGHGEP
PERTQSLSEPEFOPMARSAVS"
2954..3023
/note="Sm-4 OR SMC04506
REPEAT SM-4
predicted by Homology"
/evidence=not_experimental
complement(3036..3302)
/gene="SMC04261"
complement(3036..3302)
/gene="SMC04261"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="HYPOTHEICAL PROTEIN"
/protein_id="CAC46510.1"
/db_xref="GI:15074953"
/translation="MKFTVLEGKAAEVTAAGSVTIDAEERRIKAAQYDQWRVRSIA
TGAPIPGIQYLKMQIDVAVAGKLEOLNIPSDFTDKYWPAGI"
gene
complement(3407..4837)
/gene="gnd OR SMC04262"
complement(3407..4837)
/gene="gnd OR SMC04262"
/EC_number="1.1.1.44"

gene
/function="small molecule metabolism; energy metabolism,
carbon: oxidative branch, pentose pathway"
/note="Product confidence : probable
Gene name confidence : putative
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PROBABLY 6'-PHOSPHOGLUCONATE DEHYDROGENASE
(DECARBOXYLATING) PROTEIN"
/protein_id="CAC46511.1"
/db_xref="GI:15074954"
/translation="MSQAEIGLIGLVGWSNLAINEKGNRIAVFNRTVDATKPYA
EAGALKDIOVCETIEEPVAIRPRPITIMTKAGDPDVOOMEALKPLAKGDIIMTA
GANRRTDTRFRFDALKDSGLFTIGMGVSGEGEGRHSGTIVGCTEEYRAREVVLTS
IAAKYDSCPQVAMLENGAGHVKTKTHNGIEYADQMIAETLYGILRDLKMTAOETGE
VFGANNKRLNSYLIETEKVILKADPLTGKPIYDMLDKAGOGTQKWSVIEQNGV
VPATAIEAAVARSISSAKEEREAEEKVILGPVGEIKADRDGFIKDLNALLAARI
GAYAGGFVMAASKEFGMNLPMPTIAKIRWAGCIIRSOFLDEITTAFTKAPDANLI
VTPAFANVKESEDGALRRVYSTAVVIGLPPVALASALGYFDSYRRRGSTANVIAORD
PFCAHGFDRVDCADSHHGPWSGLNA"
5273..6691
/gene="SMC04263"
5273..6691
/gene="SMC04263"
/function="cell processes: transport of small molecules;
amino acids, amines, peptides"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE AMINO ACID CARRIER TRANSMEMBRANE
PROTEIN"
/protein_id="CAC46512.1"
/db_xref="GI:15074955"
/translation="MDTITIGLNTIFKCVLYLIGLAVGLYTVRLGFPQIVHSEFM
RVLNRSGSDAAGISPPQALWLSLRGTGMLAGVAVALYIGGGAFTFMWIAFVG
KMTAAESLADLYKIRNDGQYRGQPAEYIARGINAAPAAVITFSACILISFGVLFA
VOANSIADAVQCAFVGPRLIAGVAAVALLSGVYIFGIGIOIAFAVAVIYVPMFAAALLT
AYVLLANASAVGVLMWTIISAFGEQPAAGITGIGIAAAMGVKRLGFSNAGMGS
APNIAAVATPVPHPHSSGQFVOSIGVIFDTILICATSVMLISGLTEBGSVGTOL
TORANSTHGSAGTFTIATIFPAFTSTIGVSTAEANLTYLGGCNLMGIMNCAT
LNAVWGAVESTITVFDADADSMGLMATINLVAILLSGTIAKTRDFEOKKAGAVP
VFHAADYPELQCKIDGELWSRD"
6773..7663
/gene="SMC04264"
6773..7663
/gene="SMC04264"
/function="miscellaneous; not classified regulator"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAC46513.1"
/db_xref="GI:15074956"
/translation="MGTSCPAYDKARABERIQOENRRLRGPALAAOTSIGYRQSP
AALAEFKICLWHRMPDGAPSSIAVVPDCCSDIWSKGLAVGPDRIAPFELPAGV
TITGARFRIGAAASMLRPLREMTGQVPLHLVWGPATNIEARIKDAASPTERGLV
AALIGRLSAAEPPPDIAACVANHLOVGRSTVDDAIRTLYREIGTSEKTLRHCHHEHL
GYCAKTIGRLRIORFLIACRSRPGDTLANLVAVDAGYADQAHILAREARGLTISFTPREI
OROLARPGIGRVENRRTT"
7796..8239
gene
/function="small molecule metabolism; energy metabolism,
carbon: oxidative branch, pentose pathway"
/note="Product confidence : probable
Gene name confidence : putative
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PROBABLY 6'-PHOSPHOGLUCONATE DEHYDROGENASE
(DECARBOXYLATING) PROTEIN"
/protein_id="CAC46511.1"
/db_xref="GI:15074954"
/translation="MSQAEIGLIGLVGWSNLAINEKGNRIAVFNRTVDATKPYA
EAGALKDIOVCETIEEPVAIRPRPITIMTKAGDPDVOOMEALKPLAKGDIIMTA
GANRRTDTRFRFDALKDSGLFTIGMGVSGEGEGRHSGTIVGCTEEYRAREVVLTS
IAAKYDSCPQVAMLENGAGHVKTKTHNGIEYADQMIAETLYGILRDLKMTAOETGE
VFGANNKRLNSYLIETEKVILKADPLTGKPIYDMLDKAGOGTQKWSVIEQNGV
VPATAIEAAVARSISSAKEEREAEEKVILGPVGEIKADRDGFIKDLNALLAARI
GAYAGGFVMAASKEFGMNLPMPTIAKIRWAGCIIRSOFLDEITTAFTKAPDANLI
VTPAFANVKESEDGALRRVYSTAVVIGLPPVALASALGYFDSYRRRGSTANVIAORD
PFCAHGFDRVDCADSHHGPWSGLNA"
5273..6691
/gene="SMC04263"
5273..6691
/gene="SMC04263"
/function="cell processes: transport of small molecules;
amino acids, amines, peptides"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE AMINO ACID CARRIER TRANSMEMBRANE
PROTEIN"
/protein_id="CAC46512.1"
/db_xref="GI:15074955"
/translation="MDTITIGLNTIFKCVLYLIGLAVGLYTVRLGFPQIVHSEFM
RVLNRSGSDAAGISPPQALWLSLRGTGMLAGVAVALYIGGGAFTFMWIAFVG
KMTAAESLADLYKIRNDGQYRGQPAEYIARGINAAPAAVITFSACILISFGVLFA
VOANSIADAVQCAFVGPRLIAGVAAVALLSGVYIFGIGIOIAFAVAVIYVPMFAAALLT
AYVLLANASAVGVLMWTIISAFGEQPAAGITGIGIAAAMGVKRLGFSNAGMGS
APNIAAVATPVPHPHSSGQFVOSIGVIFDTILICATSVMLISGLTEBGSVGTOL
TORANSTHGSAGTFTIATIFPAFTSTIGVSTAEANLTYLGGCNLMGIMNCAT
LNAVWGAVESTITVFDADADSMGLMATINLVAILLSGTIAKTRDFEOKKAGAVP
VFHAADYPELQCKIDGELWSRD"
6773..7663
/gene="SMC04264"
6773..7663
/gene="SMC04264"
/function="miscellaneous; not classified regulator"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAC46513.1"
/db_xref="GI:15074956"
/translation="MGTSCPAYDKARABERIQOENRRLRGPALAAOTSIGYRQSP
AALAEFKICLWHRMPDGAPSSIAVVPDCCSDIWSKGLAVGPDRIAPFELPAGV
TITGARFRIGAAASMLRPLREMTGQVPLHLVWGPATNIEARIKDAASPTERGLV
AALIGRLSAAEPPPDIAACVANHLOVGRSTVDDAIRTLYREIGTSEKTLRHCHHEHL
GYCAKTIGRLRIORFLIACRSRPGDTLANLVAVDAGYADQAHILAREARGLTISFTPREI
OROLARPGIGRVENRRTT"
7796..8239

source	/chromosome="21" /map="21q22.3" <1.81445 /note="Accession No. AJ011929" /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="21" /map="21q22.3" /clone="p269A14, 5' partial" /clone_lib="RPC11,3-5 PAC library" <1.47014 /note="Accession No. AJ011929" /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="21" /map="21q22.3" /clone="p141D16, 5' partial" /clone_lib="RPC11,3-5 PAC library" 3304.3525 /note="MER58A" /rpt_family="DNA/MERL_type" /rpt_type=DISPERSED 4234.4579 /note="AluSx" /rpt_family="SINE/Alu" /rpt_type=DISPERSED 4693.4768 /note="L1MC/D" /rpt_family="LINE/L1" /rpt_type=DISPERSED 5206.5229 /note="(TTTTA)n" /rpt_family="Simple_repeat" /rpt_type=TANDEM complement(5230.5510) /note="Aluub" /rpt_family="SINE/Alu" /rpt_type=DISPERSED 5518.5541 /note="(T)n" /rpt_family="Simple_repeat" /rpt_type=TANDEM 5558.5989 /note="Charlie7" /rpt_family="DNA/MERL_type" /rpt_type=DISPERSED complement(5996.6135) /note="MIR" /rpt_family="SINE/MIR" /rpt_type=DISPERSED complement(6240.6541) /note="AluSx" /rpt_family="SINE/Alu" /rpt_type=DISPERSED 7101.7138 /note="(TTTA)n" /rpt_family="Simple_repeat" /rpt_type=TANDEM complement(7139.7415) /note="L1PA5" /rpt_family="LINE/L1" /rpt_type=DISPERSED complement(8130.8232) /note="L1PA5" /rpt_family="LINE/L1" /rpt_type=DISPERSED 8233.8440 /note="L1PA5" /rpt_family="LINE/L1" /rpt_type=DISPERSED 8804.9172 /note="MLT1B" /rpt_family="LTR/MaLR"	repeat_region	/rpt_type=DISPERSED 11017.11150 /note="MER5A" /rpt_family="DNA/MERL_type" /rpt_type=DISPERSED complement(11278.11746) /note="L2" /rpt_family="LINE/L2" /rpt_type=DISPERSED complement(11736.12306) /note="L2" /rpt_family="LINE/L2" /rpt_type=DISPERSED complement(12307.12591) /note="AluSg" /rpt_family="SINE/Alu" /rpt_type=DISPERSED complement(12592.12783) /note="L2" /rpt_family="LINE/L2" /rpt_type=DISPERSED 14096.14164 /note="L2" /rpt_family="LINE/L2" /rpt_type=DISPERSED 14188.14329 /note="L1MB3" /rpt_family="LINE/L1" /rpt_type=DISPERSED complement(14330.14682) /note="MLT1A1" /rpt_family="LTR/MaLR" /rpt_type=DISPERSED 14685.15070 /note="L1MB3" /rpt_family="LINE/L1" /rpt_type=DISPERSED complement(17893.18027) /note="MIR" /rpt_family="SINE/MIR" /rpt_type=DISPERSED 18438.19329 /note="MER49" /rpt_family="LTR/MER4-group" /rpt_type=DISPERSED complement(20328.20688) /note="MLT1A1" /rpt_family="LTR/MaLR" /rpt_type=DISPERSED complement(20741.20997) /note="LTR16C" /rpt_family="LTR/Retroviral" /rpt_type=DISPERSED 21744.22035 /note="AluSg" /rpt_family="SINE/Alu" /rpt_type=DISPERSED complement(22282.22412) /note="MIR" /rpt_family="SINE/MIR" /rpt_type=DISPERSED complement(22950.23031) /note="L2" /rpt_family="LINE/L2" /rpt_type=DISPERSED complement(25177.25479) /note="AluSg" /rpt_family="SINE/Alu"	repeat_region	Query Match 64.0%: Score 16; DB 9; Length 340000; Best Local Similarity 100.0%; Pred. No. 33; Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
--------	--	---------------	---	---------------	---

QY 9 gtcagtcagtcag 24
 |||
 Db 172178 CTCAGTCGAGTCAG 172193

RESULT 33

ECOBIOB 128 bp DNA ECT 03-JAN-1995
 DEFINITION Escherichia coli biotin (biob) gene, early terminator region.
 M27731
 ACCESSION M27731.1 GI:341755
 VERSION biob gene: biotin.
 KEYWORDS Escherichia coli (strain K-12) DNA.
 SOURCE Escherichia coli
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1 (bases 1 to 128)
 AUTHORS Nath, S.K.
 TITLE Attenuation of transcription of biotin genes in Escherichia coli
 JOURNAL Can. J. Microbiol. 34 (12), 1288-1296 (1988)
 MEDLINE 89167942

FEATURES
 source location/Qualifiers
 1..128
 /organism="Escherichia coli"
 /strain="K-12"
 /db_xref="taxon:562"
 gene 4..128
 /gene="biob"
 -10_signal 4..9
 /gene="biob"
 mRNA 20..>128
 /gene="biob"
 103..121
 /gene="biob"
 terminator /note="early terminator"
 28 c 34 g 35 t

BASE COUNT 31 a 28 c 34 g 35 t
 ORIGIN

Query Match 60.0%; Score 15; DB 1; Length 128;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgaatcgcagtcag 15
 |||
 Db 65 CGATCTGTCAGGT 79

RESULT 34

AF072252 636 bp mRNA PRI 28-OCT-1998
 LOCUS AF072252
 DEFINITION Homo sapiens methyl-CpG binding protein MBD2 (MBD2) gene, 3' UTR.
 AF072252
 ACCESSION AF072252
 VERSION AF072252.1 GI:3800812
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 636)
 AUTHORS Hendrich, B. and Bird, A.
 TITLE Identification and characterization of a family of mammalian
 methyl-CpG binding proteins
 Mol. Cell. Biol. 18 (11), 6538-6547 (1998)
 JOURNAL 98449942
 MEDLINE
 REFERENCE 2 (bases 1 to 636)
 AUTHORS Hendrich, B. and Bird, A.
 TITLE Direct Submission
 Submitted (15-JUN-1998) Institute of Cell and Molecular Biology,
 University of Edinburgh, Mayfield Road, Edinburgh, Scotland EH9
 3JR, UK
 FEATURES
 source location/Qualifiers
 1..636

QY 10 tcagtcagtcagtcag 24
 |||
 Db 408 TCAGTCGAGTCAG 422

RESULT 35

TCU57884 728 bp DNA PLN 30-MAR-1997
 LOCUS TCU57884
 DEFINITION Thanatephorus cucumeris strain 2R144 5.8S ribosomal RNA gene and
 internal transcribed spacers 1 and 2, complete sequence.
 U57884
 ACCESSION U57884.1 GI:1914814
 VERSION
 KEYWORDS
 SOURCE Thanatephorus cucumeris.
 ORGANISM Thanatephorus cucumeris
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes;
 Heterobasidiomycetes; Heterobasidiomycetidae; Ceratobasidiales;
 Ceratobasidiaceae; Thanatephorus.
 REFERENCE 1 (bases 1 to 728)
 AUTHORS Salazar, O., Schneider, H., Keijer, J. and Rubio, V.
 TITLE Phylogenetic relations in Rhizoctonia solani AG2 using ITS
 sequences
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 728)
 AUTHORS Rubio, V.
 TITLE Direct Submission
 Submitted (09-MAY-1996) Victor Rubio, Biotechnology Microbiana,
 Centro Nacional de Biotecnologia (CSIC-UAM), Campus
 Cantoblanco-UAM, Madrid, 28049, Spain
 FEATURES
 source location/Qualifiers
 1..728
 /organism="Thanatephorus cucumeris"
 /strain="2R144"
 /specific_host="Tulipa sp."
 /db_xref="taxon:118239"
 /tissue_type="mycelia"
 /note="isolated from tulip, from The Netherlands"
 1..294
 /note="internal transcribed spacer 1; ITS1"
 295..449
 /product="5.8S ribosomal RNA"
 450..728
 /note="internal transcribed spacer 2; ITS2"
 196 a 153 c 145 g 234 t

BASE COUNT 196 a 153 c 145 g 234 t
 ORIGIN

Query Match 60.0%; Score 15; DB 8; Length 728;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 cctcgtcagtcagtcag 19
 |||
 Db 20 CCTCGTCAGTCAG 34

SOURCE
ORGANISM
Oryza sativa DNA, clone:ALDC-a.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehretidae; Oryzaeae; Oryza.
REFERENCE
AUTHORS
1 (bases 1 to 1801)
Tsutsumi, K., Kagaya, Y., Hidaka, S., Suzuki, J., Tokairin, Y.,
Hirai, T., Hu, D.L., Ishikawa, K. and Ejiri, S.,
Structural analysis of the chloroplast and cytoplasmic
aldehyde-encoding genes implicated the occurrence of multiple loci
in rice
JOURNAL
MEDLINE
Gene 141 (2), 215-220 (1994)
94215906
REFERENCE
AUTHORS
2 (bases 1 to 1801)
Tsuchiya, T.
TITLE
Direct Submission
Submitted (28-OCT-1992) to the DDBJ/EMBL/Genbank databases. Tohru
Tsuchiya, Iwate University, Department of Agriculture, 3-18-8 Ueda,
Morioka, Iwate 020, Japan (Tel:0196-23-5171(ex.2776),
Fax:0196-25-7667)
COMMENT
Submitted (28-OCT-1992) to DDBJ by:
Tohru Tsuchiya
Institute for Cell Biology and Genetics
Faculty of Agriculture
Iwate University
3-18-8 Ueda
Morioka, Iwate 020
Japan
Phone: 0196-23-5171 x2776
Fax: 0196-24-5084.
FEATURES
source
Location/Qualifiers
1..1801
/organism="Oryza sativa"
/db_xref="taxon:4530"
/clone="ALDC-a"
TATA_signal
41..49
prim_transcript
80..1801
exon
80..179
/number=1
CDS
join(152..179,615..1663)
/codon_start=1
/product="cytoplasmic aldehyde"
/protein_id="BA02729.1"
/db_xref="GI:218157"
/translation="MSAYCGKRYKDELINAAVIGPGKGLADESTGTIGKRFASIN
VENVEDNRRAFFELFCTGALQYISGVLFTLYOKTKPKDPFVILIKAGALPGI
KVDKGTIEVAGDKEETITGQHDILKCKCKAYEAGRAFAKRAVILKIGNPSOLAD
LNAOGLAGYAIICQENGILVPIPELIVGPHDIDCAIVSEVILACIKALNEHVI
LRCGLPKMWTGSDPAKVAPEVIAETVRLQRTVPAPVAIVFLSGQSEBETIL
NLAAAKLSAKRWSLSTFSGRALQOSTILKAWAGKTEVNEKAAAEFLVCKANSEATL
GYKGDVILGEGAAESLHVKKYK"
Intron
180..614
/number=1
exon
615..1801
/number=2
BASE COUNT
418 a 521 c 439 g 423 t
ORIGIN

Query Match 60.0%; Score 15; DB 8; Length 1801;
Best local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1551 TCCTCGTCAGTGCA 1565
OY 4 tccctgcagtgca 18
|||||
|||||

RESULT 39
LOCUS HAMHP 2081 bp mRNA ROD 27-APR-1993
DEFINITION Mesocricetus auratus pregnancy protein (Hrp), complete cds.
ACCESSION M96650

VERSION
M96650.1 GI:191387
KEYWORDS
pregnancy protein; pregnancy protein 60kDa.
SOURCE
Mesocricetus auratus (strain Waterhouse) (library: lambda gt 11)
female adult, 14 days pregnant liver (maternal) cDNA to mRNA.
ORGANISM
Mesocricetus auratus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Mesocricetus.
REFERENCE
AUTHORS
1 (bases 1 to 2081)
Park, C.G.
TITLE
Cloning and sequencing of a new pregnancy marker protein from
hamster liver
JOURNAL
Thesis (1992) Biological Sciences, The Wichita State University
FEATURES
source
Location/Qualifiers
1..2081
/organism="Mesocricetus auratus"
/strain="Waterhouse"
/db_xref="taxon:10036"
/dev_stage="adult, 14 days pregnant"
/sex="female"
/tissue_type="liver (maternal)"
/tissue_id="lambda gt 11"
68..130
/gene="Hpp"
68..1330
/gene="Hpp"
/codon_start=1
/product="pregnancy protein 60 kDa"
/protein_id="AAA37078.1"
/db_xref="GI:191388"
mat_peptide
sig_peptide
CDS
68..1330
/gene="Hpp"
/product="pregnancy protein 60 kDa"
BASE COUNT
547 a 572 c 484 g 478 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 2081;
Best local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 cagggtcaggtcagc 25
|||||
|||||

DB 455 CAGGTGACAGTGACG 469
RESULT 40
LOCUS DRAJ5029 2931 bp mRNA VRT 17-DEC-1998
DEFINITION Danio rerio mRNA for Eph-like receptor tyrosine kinase rtk8.
ACCESSION AJ005029
VERSION AJ005029.1 GI:3163941
KEYWORDS
Eph-like receptor tyrosine kinase.
SOURCE
zebrafish.
ORGANISM
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE
AUTHORS
1 (bases 1 to 2931)
Cooke, J.E.
TITLE
Direct Submission
Submitted (27-MAR-1998) Cooke J.E., University College London,
Department of Anatomy and Developmental Biology, Gower Street,

REMARK London, WCIE 6BT, UK
 REVISION 2 (bases 1 to 2931)
 AUTHORS Cooke, J.E., Xu, Q., Wilson, S.W. and Holder, N.
 TITLE Characterisation of five novel zebrafish Eph-related receptor tyrosine kinases suggests roles in patterning the neural plate
 JOURNAL Dev. Genes Evol. 206, 515-531 (1997)
 REFERENCE 3 (bases 1 to 2931)
 AUTHORS Shiom, K.
 TITLE Direct Submission
 JOURNAL Submitted (26-MAY-1998) Shiom, K., University College London, Department of Anatomy and Developmental Biology, Gower Street, London, WC1E 6BT, UK
 REFERENCE 4 (bases 1 to 2931)
 AUTHORS Durbin, L., Brennan, C., Shiom, K., Cooke, J., Barrios, A., Shannugalingam, S., Guthrie, B., Lindberg, R. and Holder, N.
 TITLE Eph signalling is required for segmentation and differentiation of the somites
 JOURNAL Genes & development, 12 (19), 3096-3109 (1998)
 MEDLINE 98438455
 PUBMED 9765210
 COMMENT On May 28, 1998 this sequence version replaced gi:3005906.
 FEATURES
 source Location/Qualifiers
 1..2931
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 1..2931
 /codon_start=1
 /product="Eph-like receptor tyrosine kinase rtk8"
 /protein_id="CAA06302.1"
 /db_xref="GI:3163942"
 /db_xref="SPTREMBL:O73878"
 /translation="MDRVCIMALSNPFWSTGLVASPEFLMNTKLETSDLRWITYP
 SCPPBEMSGDLDEGNSVPTPOVCPMDSSVSNMLTRFLPRGASQVVEIRTYME
 CSMPASFRTCKEFTNLYYQSEDTASATHPAMENPYSKVDTVADFLRKGGERK
 SNKTVAVGPLSLFEGYLAFOQTGACMALISVVFPRKCPAVSRAPSSPELTPSLV
 QQAEGYVDNSAPTGCTAPPTMFCGEGDQWGPSPSTCAKPGYPPVSDRCRGL
 GYKASVGSGLCVRCPDNTSHSAGSLCVRGYRGLVSDLPDSACTKPPSPRSII
 YOINDVYVTEEMSEPLDRGRSLSTVSECMHGRGLCYOCAGSTIYRPGCMQVSR
 VITRGLPHRTTYFYLAONGVSAVSHTSASSVNTTISRDAVAVYSGIRIKAES
 SVSISWTPPOTQHSIDYOLRLSLKGQDQGWYVSSRSSVVLNLSRSQYQVVR
 ARTAAGVGHSSASVISTLPDDSPSRMLTGVLAIGLLIAVAVAFVCFRST
 RRDDPPDKSGQFLMGQGLKIVYIDPFYEDPNVAREFAKEIDVSFKIEEVGAEF
 GEYCRGLKVPKKNVVAIKTLKGDTOKRDRPSEASIMGCFQHPNIILHEGYT
 ASCPVMITPEYMNGALDSFLRLINDGQFPIOLVGMRLGASCMKYLSPSPYRRLA
 ARNLVNSNLVCKVSDPGLSRFLTENSDDPTTSSLGKPIPIWTPAPELARPKFTSA
 SDWASVGIWMEVMSFEERYWDSMODVINAIEDYRLLPPEPCPASLHQLMLDCWQ
 KERSSRPFCALVSALDRILRNPSALIKITCRIPDGSHPLLDORAPPLSHCSSVADW
 LRAIKMERYEDAFMQAGFTAIQIHITHTEDLRLIGVTLAGHOKILLSSVQTLRHGG
 SLRY"

BASE COUNT 602 a 847 c 852 g 630 t
 ORIGIN

Query Match 60.0%; Score 15; DB 5; Length 2931;
 Best Local Similarity 100.0%; Pred. No. 2, 1e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 tcaggtcaggtcag 24
 |||||

Db 1509 TCAGGTGACAGTCAG 1523

RESULT 41
 HSMETHYL7
 LOCUS HSMETHYL7 3829 bp DNA PRI 28-SEP-1999
 DEFINITION Homo sapiens methyl-CpG binding protein 2 (MBD2) gene, exon 7.
 ACCESSION AF120994
 VERSION AF120994.1 GI:5929753
 KEYWORDS
 SEGMENT 7 of 7
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 3829)
 AUTHORS Hendrich, B., Abbott, C., McQueen, H., Chambers, D., Cross, S. and Bird, A.
 TITLE Genomic structure and chromosomal mapping of the murine and human mbd1, mbd2, mbd3, and mbd4 genes
 JOURNAL Mammalian Genome 10 (9), 906-912 (1999)
 MEDLINE 99373255
 PUBMED 10441743
 REFERENCE 2 (bases 1 to 3829)
 AUTHORS Hendrich, B. and Bird, A.
 TITLE Direct Submission
 JOURNAL Submitted (19-JAN-1999) ICMB, University of Edinburgh, Darwin Building, King's Buildings, Edinburgh, Scotland EH9 3JR, United Kingdom
 FEATURES
 source Location/Qualifiers
 1..3829
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18q21"
 /order(AFI20988,1:253,1993,AF120989,1:1,4465,AF120990,1:1,2374,AF120991,1:1,429,AF120992,1:1,682,AF120993,1:1,742,1,1340)
 /gene="MBD2"
 230,1340
 /gene="MBD2"
 /note="MBD2"
 /note="alternative: alternate polyadenylation site"
 /number=7
 230,704
 /gene="MBD2"
 /note="alternative: major polyadenylation site"
 /number=7

exon

BASE COUNT 1041 a 801 c 716 g 1270 t 1 others
 ORIGIN

Query Match 60.0%; Score 15; DB 9; Length 3829;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 tcaggtcaggtcag 24
 |||||

Db 1112 TCAGGTGACAGTCAG 1126

RESULT 42
 LOCUS HSA317970 4380 bp mRNA PRI 05-MAY-2001
 DEFINITION Homo sapiens mRNA for ataxin-2 related domain protein (A2D-A gene).
 ACCESSION AJ317970
 VERSION AJ317970.1 GI:13992552
 KEYWORDS A2D-A gene; ataxin-2 related domain protein.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 4380)
 AUTHORS Meunier, C.F.
 TITLE Cloning and characterisation of family of proteins associated with

MP1
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 4380)
 AUTHORS Courtois, G.M.
 TITLE Direct Submission
 JOURNAL Submitted (30-APR-2001) Courtois G.M., Hematologie, INSERM U363
 Hospital Cochin, 27 rue du Faubourg Saint Jacques PARIS, 75014, FRANCE

FEATURES
 source Location/Qualifiers
 ..4380
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

```

/chrosome="16"
/cell_line="U7 and Jurkat"
169. .3396
/gene="A2D-A"
169. .3396
/gene="A2D-A"
/codon_start=1
/pseudo
/function="unknown"
/evidence=experimental
/product="ataxin-2 related domain protein"
BASE COUNT      925 a 1476 c 1143 g 836 t
ORIGIN
Query Match      60.0%; Score 15; DB 9; Length 4380;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 caggtgcaggtcagc 25
Db 3447 CAGGTGCAGGTGAGC 3433

RESULT 43
AF034373 4674 bp mRNA PRI 01-NOV-1998
LOCUS AF034373
DEFINITION Homo sapiens ataxin-2-like protein A2LP (A2LG) mRNA, complete cds.
ACCESSION AF034373
VERSION AF034373.1 GI:3820483
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 4674)
AUTHORS Xia,J.-h., Liu,C.-Y., Wang,D.-a., Ruan,Q.-g. and Deng,H.-x.
TITLE A splicing form of human ataxin-2 like gene obtained from adult
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 4674)
TITLE Xia,J.-h., Liu,C.-Y., Wang,D.-a., Ruan,Q.-g. and Deng,H.-x.
JOURNAL Direct Submission
SUBMITTED (12-NOV-1997) National Lab of Medical Genetics of China,
Humam Medical University, 88 Xiangya Road, Changsha, Hunan 410078,
P. R. China
FEATURES
Source Location/Qualifiers
1..4674
/organism="Homo sapiens"
/db_xref="taxon:9606"
/def_stage="adult"
/tissue_type="brain"
1..4674
/gene="A2LG"
63..3218
/codon_start=1
/product="ataxin-2-like protein A2LP"
/db_xref="GI:3820484"
/translation="MGWRKGRDAHPPPAPRPPPLPPKGAASSRKGSLSPPPRGSPP
P/TAASGPIPLASRALQGPSPPLSPILSNSSPCRHVVEASAATLAPPAAP
HATCRGSPARSHSPNGGLPGPLATSAAPGPAAAPCLGVAAGSLRAGAGT
LAPPPPOHOERPPGAATGASRGOSTGCKGPPSPVFEQVYNSRMLHRTLVAGST
CDVKNKGTVEGTEKLSKFELEDAVVRKASPPAGGRREDYDTWYFKSDML
VHRNVDENATADKFTDLSAIVANSKYNGBKEKVLQREWGDSDDYLEDKMSNG
WDPEWKEEENGVKTYDSSLSSTYPLEKDNSEEFQRELRAAQLARETSSPO
YRLRIAMENDGRTEEKHSVAVORSGRSPGLASREGKYLPLPQREVGPGVRC
SSRKGRCGLSLPPRPHLDNNSPQSGSEARGLNGSPRMSKQAPRLRGAKTLS
PSRLEENFSTSSGSDVSPRSPSAAPAPISASCPPEPPIGSAVPTSSASIPVTS
VSDPGVSTISPAKPTSLATPDVKELESTKRGTELEPROELRIAGVPCIONKREO
LELRKGAQFKLQPSSESPNSLDPPFPRLKEPPKGEKREVDGLTSEPMGSPVSSK

```

```

TESVSDKEDKPLASGSGTEGPEPPPCPSQTSPPVGLIKGDKGEGVAEYKKS
TLPNKKERNPTKPLLSVSKSTSTSTSPGRHSPSLPVLTAQSGSLTSPOTISYIP
QIHMGPAVQADQMPYPPVSNVSGQGYRKAKGSLPPQRSDQHPQASAPPMQAAA
AGPLVLAATPVSSYIPVNDQPPGQPMAMQAHYPSQVPAFPMLOSMPRLTSGSH
QAVISSSTPQYSAEQPTPOALYATVHOSYHHATOLHAHOPQATPTTGSQSOHA
APSPVOVPMAGGAEWSCRCNGWPEEGTELGYSIEMRGASGASSELLACVALNPLPSSIR
RGPRHMAVDSSHRICTTQGP"
BASE COUNT      971 a 1566 c 1244 g 893 t
ORIGIN
Query Match      60.0%; Score 15; DB 9; Length 4674;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 caggtgcaggtcagc 25
Db 3741 CAGGTGCAGGTGAGC 3727

RESULT 44
AB037816/c
LOCUS AB037816 4886 bp mRNA PRI 14-MAR-2000
DEFINITION Homo sapiens mRNA for KIAA1395 protein, partial cds.
ACCESSION AB037816
VERSION AB037816.1 GI:7243170
KEYWORDS
SOURCE Homo sapiens brain cDNA to mRNA, clone.lib:pbuescriptII SK plus
clone:hj07927.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (sites)
AUTHORS Nagase,T., Kikuno,R., Ishikawa,K.I., Hirose,M. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes.
XVI. The complete sequences of 150 new cDNA clones from brain which
code for large proteins in vitro
JOURNAL DNA Res. 7 (1), 65-73 (2000)
AUTHORS 2 (bases 1 to 4886)
TITLE Ohara,O., Nagase,T. and Kikuno,R.
JOURNAL Direct Submission
SUBMITTED (31-JAN-2000) to the DDBJ/EMBL/GenBank databases. Osamu
Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology;
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(Email:cdna@kazusa.or.jp URL:http://www.kazusa.or.jp/huge/,
Tel:+81-438-52-3913, Fax:+81-438-52-3914)
FEATURES
Source Location/Qualifiers
1..4886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="pbuescriptII SK plus"
/clone_lib="pbuescriptII SK plus"
/tissue_type="brain"
1..4884
/gene="KIAA1395"
<1..>4884
/gene="KIAA1395"
/feature="Start codon is not identified."
/translation="DARMASSERRARPAKINRTVAEVRKQVRSRSGSSRSG
SLGKPLTEVERPDEPVLVSRPPAPRGLRDLVEPRADDELLQRPCCRTTSG
IPKDKLDAQVRAAEVMTIEDVAVIHRKTYTSLAIVTDTDTREKQGLPQVDFD
DASGERGSGSDNSRSGSPEDTPRSGASSIFDLRLAASLPLSLERAAPD
VDRNETLRQHRPDPALTLTPADDEDEAVRCSRPPRPPRFGORILLVCLSKFEI
ELPFTGLIADYVREKKISENFPDLSMSMGGLRAHCTHPATILARSAIFSVT
YSPPIPLVITLKVLEVLQGDISECCPEYMLKEDVTAANKREKLEKLLAASQFTSG
RYRMEFAVAVHIANIVSAGOLDRDSGSEERPAWRDRRRRQPORASSGDACFP
SGFRATITVITNFKQERLSDDEDLRFLADMRPSSLRLRLRPVLAQKIDISPAF
ENPHCLSPELLHTRKPYPPDPRGPTKTELEPPAREVYAPHTSYNNLLVYVPSLNFSS

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:17:06 ; Search time 368.09 Seconds
(without alignments)
58.228 Million cell updates/sec

Title: US-09-396-196f-9

Perfect score: 25

Sequence: 1 cgatcctcgtcaggtcgcaggtcagc 25

Scoring table: OLIGO-MUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 9

Total number of hits satisfying chosen parameters: 44981

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

N_Geneseq_1101.*
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
1	25	100.0	839	22	AA162941		Human genomic DNA
2	25	100.0	1041	20	AA101303		E. coli biotin syn
3	25	100.0	1084	10	AA191329		E. coli Bio B gene.
4	25	100.0	1121	7	AA160496		Sequence encoding
5	25	100.0	5872	15	AA062386		Biotin-biosynthet
6	16	64.0	450	22	AA112043		Probe #1976 for ge
7	16	64.0	450	22	AA113378		Probe #12064 used t
8	16	64.0	450	22	AA101968		Probe #1959 used t
9	15	60.0	34	16	AA075832		Sense primer to am
10	15	60.0	353	21	AA056641		Eucalyptus grandis
11	15	60.0	586	21	AA075553		Human ORFX ORF1118

12	C	12	60.0	4674	20	AA090421	Human ataxin-2 lik
13	C	13	60.0	6372	22	AA143851	Human CLASP-7 enco
14	C	14	56.0	327	21	AA000048	Human secreted pro
15	C	15	56.0	363	22	AA142390	Nucleotide sequenc
16	C	16	56.0	370	22	AA134737	Human colon cancer
17	C	17	56.0	374	22	AA164981	Novel human polyn
18	C	18	56.0	423	14	AA045597	Sequence encoding
19	C	19	56.0	423	20	AA239427	Antibody B17X2 VH
20	C	20	56.0	423	20	AA223965	Human B17X2 antibo
21	C	21	56.0	444	18	AA180181	Monoclonal antibody
22	C	22	56.0	447	20	AA220406	IgG antibody 2.4.4
23	C	23	56.0	613	16	AA078943	Human immunoglobul
24	C	24	56.0	630	16	AA078948	Human immunoglobul
25	C	25	56.0	981	21	AA115905	Human protein clon
26	C	26	56.0	1100	21	AA115915	Human protein clon
27	C	27	56.0	1130	21	AA139064	Human secreted pro
28	C	28	56.0	1135	21	AA139087	Human secreted pro
29	C	29	56.0	1160	21	AA129322	Glycine max iso fla
30	C	30	56.0	1161	22	AA191019	Human secreted pro
31	C	31	56.0	1173	21	AA298136	Human signal pepti
32	C	32	56.0	1186	21	AA098064	Human colon cancer
33	C	33	56.0	1186	22	AA135018	Human colon cancer
34	C	34	56.0	1191	19	AA198726	DNA encoding a S.
35	C	35	56.0	1368	22	AA144623	Novel protein kin
36	C	36	56.0	1634	19	AA296242	S. pneumoniae deri
37	C	37	56.0	1634	19	AA137360	Streptococcus pneu
38	C	38	56.0	1758	19	AA144597	Mouse uncoupling p
39	C	39	56.0	1814	22	AA159259	Human polynucleoti
40	C	40	56.0	1816	22	AA161045	Human polynucleoti
41	C	41	56.0	2014	22	AA161045	Human DNA associat
42	C	42	56.0	3720	21	AA236323	Mechanical stress
43	C	43	56.0	4027	21	AA107382	Human DNA associat
44	C	44	56.0	7215	22	AA107373	Human cDNA encodin
45	C	45	56.0	37856	21	AA111992	S. cellulosum DNA
46	C	46	56.0	66685	22	AA107380	Human genomic DNA
47	C	47	56.0	273254	21	AA181914	Chlamydia pneumoni
48	C	48	56.0	534720	19	AA130458	Rhizobium species
49	C	49	56.0	536165	19	AA130459	Rhizobium species
50	C	50	52.0	166	20	AA119643	Complement system
51	C	51	52.0	123	20	AA185925	Oligonucleotide us
52	C	52	52.0	269	22	AA122207	Probe #12140 for g
53	C	53	52.0	276	21	AA100049	Human secreted pro
54	C	54	52.0	279	22	AA126888	Probe #16821 for g
55	C	55	52.0	279	22	AA155684	Probe #24370 used
56	C	56	52.0	281	22	AA143857	Human CLASP-7 Intr
57	C	57	52.0	303	22	AA165933	C glutamylcystine
58	C	58	52.0	351	21	AA108252	Human secreted pro
59	C	59	52.0	363	20	AA185929	DNA encoding the h
60	C	60	52.0	363	20	AA179527	Heavy chain coding
61	C	61	52.0	375	21	AA122004	Human breast and o
62	C	62	52.0	375	21	AA101754	Human secreted pro
63	C	63	52.0	378	18	AA161237	Human anti-RSV mon
64	C	64	52.0	378	18	AA161239	Human anti-RSV mon
65	C	65	52.0	395	22	AA112969	Probe #2902 for ge
66	C	66	52.0	400	21	AA100033	Human secreted pro
67	C	67	52.0	423	20	AA179207	Anti-HIV-1 gp120 a
68	C	68	52.0	437	21	AA142289	Human 5' Est isola
69	C	69	52.0	463	22	AA116085	Probe #6018 for ge
70	C	70	52.0	463	22	AA1138303	Probe #6989 used t
71	C	71	52.0	472	21	AA175419	Human ORFX ORF974
72	C	72	52.0	472	21	AA120408	IgG antibody 2.3.2
73	C	73	52.0	480	22	AA165984	C glutamylcystine
74	C	74	52.0	527	20	AA187894	EST clone FC380.
75	C	75	52.0	529	22	AA113160	Human cDNA clone (
76	C	76	52.0	546	16	AA078943	Human immunoglobul
77	C	77	52.0	553	22	AA197949	Murine 7-transemb
78	C	78	52.0	588	22	AA117685	Probe #7618 for ge
79	C	79	52.0	588	22	AA142613	Probe #11305 used
80	C	80	52.0	593	22	AA140100	Probe #8786 used t
81	C	81	52.0	623	21	AA075197	Human ORFX ORF752
82	C	82	52.0	732	21	AA075028	Human cervical can
83	C	83	52.0	750	20	AA158277	zee may's SCBR pro
84	C	84	52.0	879	16	AA101341	Human FK506 bindin

C 85	13	52.0	880	21	AAE20389	Human FK-506 bindi
C 86	13	52.0	880	21	AA8346027	Human adenosine re
C 87	13	52.0	911	22	AAH29728	S cerevisiae apopta
C 88	13	52.0	933	22	AAH32118	Human Olfactory re
C 89	13	52.0	939	22	AAH67432	C glutamicum codin
C 90	13	52.0	952	22	AAH32030	Human olfactory re
C 91	13	52.0	963	22	AAH67430	C glutamicum codin
C 92	13	52.0	1040	20	AAH56725	Zea mays SCLBr pro
C 93	13	52.0	1081	22	AAF90596	Cotton transcript
C 94	13	52.0	1101	19	AAAF3298	Human secreted pro
C 95	13	52.0	1113	22	AAF71611	Corynebacterium gl
C 96	13	52.0	1121	22	AAAD07933	Human TPC protein
C 97	13	52.0	1173	22	AAAC68663	Human DIB2 cDNA.
C 98	13	52.0	1215	22	AAH45073	Murine FM-3 coding
C 99	13	52.0	1323	18	AAH87005	DNA encoding S. ce
C 100	13	52.0	1341	20	AAH57178	Human KDN33 intron

ALIGNMENTS

RESULT	1
AAI62941	
ID	AAI62941 standard; DNA; 839 BP.
XX	
AC	AAI62941;
XX	
DT	22-OCT-2001 (first entry)
XX	
DE	Human genomic DNA SEQ ID NO 269.
XX	
KW	Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;
KW	immunosuppressive; antiinflammatory; anti-HIV; antibacterial; antiviral;
KW	antiparkinsonian; antispasmodic; antianemic; antiarthritic; cancer;
KW	antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW	antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW	antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW	neurological disease; infection; nephrotropic; gene therapy; vaccine;
ds.	
XX	
OS	Homo sapiens.
XX	
PN	WO20015449-A1.
XX	
PD	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01346.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUL-2000; 2000US-0216880.
PR	14-JUL-2000; 2000US-0218290.
PR	14-AUG-2000; 2000US-0225447.
PR	01-SEP-2000; 2000US-0229343.
PR	06-SEP-2000; 2000US-0230437.
PR	08-SEP-2000; 2000US-0231243.
PR	25-SEP-2000; 2000US-0234997.
PR	29-SEP-2000; 2000US-0236367.
PR	13-OCT-2000; 2000US-0239957.
PR	08-NOV-2000; 2000US-0246476.
PR	08-NOV-2000; 2000US-0246477.
PR	08-NOV-2000; 2000US-0246525.
PR	08-NOV-2000; 2000US-0246526.
PR	08-NOV-2000; 2000US-0246528.
PR	17-NOV-2000; 2000US-0249210.
PR	17-NOV-2000; 2000US-0249211.
PR	17-NOV-2000; 2000US-0249214.
PR	17-NOV-2000; 2000US-0249265.
PR	01-DEC-2000; 2000US-0250160.
PR	01-DEC-2000; 2000US-0250381.
PR	05-DEC-2000; 2000US-0251030.
PR	05-DEC-2000; 2000US-0251988.

PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI: 2001-476225/51.
 DR
 XX Novel plasma membrane associated proteins useful for diagnosing,
 PT treating, preventing and/or prognosing disorders related to the
 PT proteins, including cancer, immune response and neuronal disorders -
 XX
 PS Example 2; SEQ ID NO 269; 532zp + Sequence Listing; English.

CC The invention relates to novel genes (AA162752-AA162961) and proteins
CC (AA042347-AA042345) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification
CC The nucleic acids, proteins, antibodies and (an)tagonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus and
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at http://wipo.int/pub/published_pat_sequences.
XX

Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other;

Query Match	100.0%	Score: 25	DB: 23	Length: 839
Best Local Similarity	100.0%	Pred. No.	0	0002
Matches: 25	Conservative: 0	Mismatches: 0	Indels: 0	Gaps: 0
QY	1	cgatcctcgatcgatcgatcgatcgc	25	
db	137	cgatcctcgatcgatcgatcgatcgc	161	

RESULT	2
AAx01303	
ID	AAx01303 standard; DNA; 1041 BP.
XX	
AC	AAx01303;
XX	
DT	12-APR-1999 (first entry)
XX	
DE	E. coli biotin synthetase (BioB) coding sequence.
XX	
KW	DAP aminotransferase; diaminopelargonic acid; transgenic plant;
XX	biotin synthase; biotin production; vitamin H; BioB; ss.
OS	Escherichia coli.
XX	
PN	US5869719-A.
XX	
PD	09-FEB-1999.
XX	
PF	30-APR-1997; 97US-0846338.
XX	
PR	30-APR-1997; 97US-0846338.
XX	
PR	08-MAR-1995; 95US-0401068.
XX	

PA (NOVS) NOVARTIS FINANCE CORP.
 XX Patton DA:
 XX WPI: 1999-152902/13.
 DR P-PSDB; AAW73906.
 XX
 PT Transgenic plants with high biotin levels - transformed with DNA
 encoding di:amino-pelargonic acid amino-transferase or biotin
 synthase
 XX
 PS Example 2: Column 37-40: 34pp; English.
 XX
 CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
 be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.
 CC
 XX Sequence 1041 BP: 262 A; 273 C; 305 G; 201 T; 0 other:
 SO

Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.0002;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctcgtcaggtcagtcagc 25
 ||||||||||||||||||||
 Db 105 cgatcctcgtcaggtcagtcagc 129

RESULT 3
 AAN91329
 ID AAN91329 standard; DNA: 1084 BP.
 XX
 AC AAN91329;
 XX
 DT 15-FEB-1990 (first entry)
 XX
 DE E.coli Bio B gene.
 XX
 KM E.coli: Bio B gene; biotin.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT CDS 24..1064
 FT /*tag-a
 XX
 PN GB2216530-A.
 XX
 PD 11-OCT-1989.
 XX
 PF 17-MAR-1989; 89GB-0006210.
 XX
 PR 22-MAR-1988; 88GB-0006804.
 PR 17-MAR-1989; 89GB-0006210.
 XX
 PA (UKMG-) UK MIN. AGRIC. FISH.
 XX
 PI Pearson BM, McKee RA;
 XX
 DR WPI: 1989-295085/41. P-PSDB P91392
 XX
 PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.
 XX
 PS Table 3; page 33-4; 52pp; English.
 XX
 CC The gene can be used in a plasmid for expression of enzymes of the biotin

CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
 CC Lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.
 CC
 XX Sequence 1084 BP: 271 A; 286 C; 318 G; 209 T; 0 other:
 SO

Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 0.0002;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctcgtcaggtcagtcagc 25
 ||||||||||||||||||||
 Db 128 cgatcctcgtcaggtcagtcagc 152

RESULT 4
 AAN60496
 ID AAN60496 standard; DNA: 1121 BP.
 XX
 AC AAN60496;
 XX
 DT 17-OCT-1991 (first entry)
 XX
 DE Sequence encoding biotin synthesising enzyme.
 XX
 KM Biotin synthetase enzyme; E.coli; desthiobiotin; ds.
 XX
 FH Key Location/Qualifiers
 FT CDS 42..1082
 FT /*tag-a
 XX
 PN JP61149091-A.
 XX
 PD 07-JUL-1986.
 XX
 PF 24-DEC-1984; 84JP-0272605.
 XX
 PR 24-DEC-1984; 84JP-0272605.
 XX
 PA (NIPS) NIIPPON SODA KK.
 XX
 DR WPI: 1986-216622/33.
 DR P-PSDB; AAP60536.
 XX
 PT Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
 XX
 PS Disclosure; Page 534; 23pp; Japanese.
 XX
 CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing desthiobiotin.
 XX
 XX Sequence 1121 BP: 290 A; 301 C; 319 G; 211 T; 0 other:
 SO

Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 0.0002;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctcgtcaggtcagtcagc 25
 ||||||||||||||||||||
 Db 146 cgatcctcgtcaggtcagtcagc 170

RESULT 5
 AA062386
 ID AA062386 standard; DNA: 5872 BP.
 XX

```

AC  AAG52386;
XX
XX  16-NOV-1994 (first entry)
DT
XX
XX  Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
DE
XX
XX  Biotin; expression; enterobacteria; vitamin H; synthesis;
XX  plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
XX  promoter; plac; biotin synthase; KAPA synthase;
XX  8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
XX  dehydrobiotin synthase; DAPA synthase;
XX  S-adenosyl-L-methionine; 8-amino-7-oxononanoate aminotransferase;
XX  seborrhoea; dermatitis; ds.
XX
XX  Escherichia coli DSM498.
OS
XX
XX  Key
FH  Location/Qualifiers
FT  1..96
FT  promoter
FT  /tag= a
FT  /function= "promoter plac"
FT  /evidence= EXPERIMENTAL
FT  23..28
FT  -35_signal
FT  /tag= b
FT  /standard_name= "promoter plac"
FT  45..50
FT  -10_signal
FT  /tag= c
FT  /evidence= EXPERIMENTAL
FT  /standard_name= "promoter plac"
FT  105..109
FT  RBS
FT  /tag= d
FT  /evidence= EXPERIMENTAL
FT  /standard_name= "bioB RBS no. 9"
FT  117..1157
FT  CDS
FT  /tag= e
FT  /product= "biotin synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "bioB"
FT  /number= 1
FT  1141..1146
FT  RBS
FT  /tag= f
FT  /standard_name= "bioF RBS"
FT  1154..2311
FT  CDS
FT  /tag= g
FT  /EC_number= 2.3.1.47
FT  /product= "KAPA synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "bioF"
FT  /number= 2
FT  /standard_name= "8-amino-7-oxononanoate synthase"
FT  2284..2288
FT  RBS
FT  /tag= h
FT  /standard_name= "bioC RBS"
FT  2295..3050
FT  CDS
FT  /tag= i
FT  /function= "involved in pimeloyl-CoA synthesis"
FT  /product= "protein"
FT  /gene= "bioC"
FT  /number= 3
FT  3030..3033
FT  RBS
FT  /tag= j
FT  /standard_name= "bioD RBS"
FT  3043..3753
FT  CDS
FT  /tag= k
FT  /EC_number= 6.3.3.3
FT  /product= "DTB synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "bioD15"
FT  /number= 4
FT  /standard_name= "dehydrobiotin synthase"
FT  3712..3750
FT  misc_RNA
FT  /tag= l
FT  /note= "bioD15 substitution"
FT  3742..3746
FT  RBS

```

```

FT  /tag= m
FT  /standard_name= "bioA RBS"
FT  3750..5039
FT  CDS
FT  /tag= n
FT  /EC_number= 2.6.1.62
FT  /product= "DAPA synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "bioA"
FT  /number= 5
FT  /standard_name= "S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase"
FT  5088..5093
FT  RBS
FT  /tag= o
FT  /standard_name= "ORF1 RBS"
FT  5098..5574
FT  CDS
FT  /tag= p
FT  /function= "unknown, involved in biotin synthesis"
FT  /product= "protein"
FT  /evidence= EXPERIMENTAL
FT  /gene= "ORF1"
FT  /number= 6
FT  terminator
FT  5583..5644
FT  /tag= q
FT  /standard_name= "rho-independent transcriptional terminator"
FT  5583..5605
FT  stem_loop
FT  /tag= r
FT  5583..5605
FT  WO9408023-A.
FT  PD  14-APR-1994.
FT  PE  01-OCT-1993; 93WO-EP02688.
FT  PR  02-OCT-1992; 92CH-0003124.
FT  PR  15-JUL-1993; 93CH-0002134.
FT  PA  (LONZ ) LONZA AG.
FT  PI  Birch O, Brass J, Fuhrmann M, Shaw N;
FT  DR  WPI: 1994-135587/16.
FT  DR  P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
FT  PT  Biotechnological biotin prodn. using enterobacterial biotin-gene
FT  - providing vitamin H in high yield
FT  PS  Claim 1; Fig 6, page 47-55 and 60-65; 92pp; German.
FT  XX
FT  CC  The sequence is derived from plasmid pB030A-15/9 contg. the
FT  CC  bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
FT  CC  of biotin, arranged in a transcription unit. Microorganisms
FT  CC  contg. these DNA fragments or plasmids may be used in the prodn.
FT  CC  of biotin. Biotin (vitamin H) may prevent seborrhoea, dermatitis,
FT  CC  loss of appetite and tiredness.
FT  XX
FT  SO  Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;
FT
FT  Query Match 100.0%; Score 25; DB 15; Length 5872;
FT  Best Local Similarity 100.0%; Pred. No. 0.00017;
FT  Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
FT
FT  OY 1 cgatcctgcagtcgagtcagc 25
FT  |||||||||||||||||||
FT  Db 221 cgatcctgcagtcgagtcagc 245
FT
FT  RESULT 6
FT  AAI12043/c
FT  ID AAI12043 standard; DNA: 450 BP.
FT  XX
FT  AC AAI12043;

```

```

XX 12-OCT-2001 (first entry)
XX
XX
XX Probe #1976 for gene expression analysis in human cervical cell sample.
XX
XX Probe: human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer; ss.
XX
XX Homo sapiens.
XX
XX W0200157278-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 25; SEQ ID No 1976; 487bp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENPs). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging
XX of diseases of the cervix, notably cervical cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 450 BP; 121 A; 117 C; 124 G; 88 T; 0 other;
XX
XX
XX Query Match 64.0%; Score 16; DB 22; Length 450;
XX Best Local Similarity 100.0%; Pred. No. 9.4;
XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 9 gtcagtgacagtcag 24
XX ||||||||||||
XX 423 gtcagtgacagtcag 408
XX
XX
XX RESULT 7
XX ID AAI33378/C
XX ID AAI33378 standard; DNA; 450 BP.
XX
XX AC AAI33378;
XX
XX DT 17-OCT-2001 (first entry)
XX
XX DE Probe #2064 used to measure gene expression in human placenta sample.
XX
XX Probe: microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX

```

```

PN W0200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-48897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID No 2064; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 450 BP; 121 A; 117 C; 124 G; 88 T; 0 other;
XX
XX
XX Query Match 64.0%; Score 16; DB 22; Length 450;
XX Best Local Similarity 100.0%; Pred. No. 9.4;
XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 9 gtcagtgacagtcag 24
XX ||||||||||||
XX 423 gtcagtgacagtcag 408
XX
XX
XX RESULT 8
XX ID AAI01968/C
XX ID AAI01968 standard; DNA; 450 BP.
XX
XX AC AAI01968;
XX
XX DT 09-OCT-2001 (first entry)
XX
XX DE Probe #1959 used to measure gene expression in human breast sample.
XX
XX Probe: human; breast disease; breast cancer; development disorder; ss;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
XX
XX W0200157270-A2.
XX
XX 09-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US00661.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX

```

XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-476286/51.
 XX Novel single exon nucleic acid probe used to measuring gene expression
 PT in a human breast -
 XX
 PS Claim 25; SEQ ID No 1959; 322pp; English.
 CC The present invention relates to novel single exon nucleic acid probes.
 CC The present sequence is one such probe. The probes are useful for
 CC measuring human gene expression in a human breast sample, where the probe
 CC hybridises at high stringency to a nucleic acid expressed in the human
 CC breast. The probes are useful for predicting, diagnosing, grading,
 CC staging, monitoring and prognosing diseases of the human breast,
 CC particularly those diseases with polygenic aetiology. The diseases
 CC include: breast cancer, disorders of development, inflammatory diseases
 CC of the breast, fibrocystic changes, proliferative breast disease and
 CC non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 CC
 SQ Sequence 450 BP; 121 A; 117 G; 124 G; 88 T; 0 other;

Query Match 64.0%; Score 16; DB 22; Length 450;
 Best Local Similarity 100.0%; Pred. No. 9.4;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 gtcaggtcaggtcag 24
 |||||
 Db 423 GTCAGGTGACGTGAG 408

RESULT 9
 AA075832
 ID AA075832 standard; DNA: 34 BP.
 XX
 AC AA075832;

DT 18-AUG-1995 (first entry)

XX Sense primer to amplify Non-A Non-B hepatitis virus for analysis.

XX Non-A Non-B hepatitis virus; structural region; CDNA to genomic RNA;
 KW detection; reagent; anti-Non-A Non-B hepatitis virus antibody;
 KW vaccine; antigen; epitope; diagnosis; ss.

OS Synthetic.

PN EP628572-A.

PD 14-DEC-1994.

PF 27-MAY-1994; 94EP-0108256.

XX 28-MAY-1993; 93JP-0126709.

PR 02-MAR-1994; 94JP-0032201.

XX (ARIMA) ARIMA T.

PA (EISA) EISA CO LTD.

PI Aoyama M, Arima T, Hosoda T, Iwasaki Y, Obara T;

PI Sawada T, Tohmatsu J;

XX WPI; 1995-015655/03.

PT New non-A non-B hepatitis virus sub-type - used to develop prods.
 PT for detection, diagnosis, prevention and treatment of non-A non-B
 PT hepatitis.

PS Example 2; Page 49; 59pp; English.
 CC This primer is based on nucleotides 778-800 of AA075818 (a part of
 CC the Non-A Non-B hepatitis virus genome encoding the non-structural
 CC protein). It is used in conjunction with AA075833 to amplify nucleotides
 CC 800-1326 of AA075818. The nucleotide sequences (see also AA075817-19)
 CC were isolated from the plasma of donors in Japan with high s-GTP levels,
 CC and were found to be different from previously reported NANB hepatitis
 CC viruses. The DNA can be used as a reagent for detecting the NANB
 CC hepatitis viral gene. The polypeptides can be used as reagents for
 CC detecting anti-NANB hepatitis antibodies or as a NANB hepatitis viral
 CC vaccine.
 SQ Sequence 34 BP; 6 A; 9 C; 12 G; 7 T; 0 other;

Query Match 60.0%; Score 15; DB 16; Length 34;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gatctctcaggtc 16
 |||||
 Db 8 gatctctcaggtc 22

RESULT 10
 AAC56641/C
 ID AAC56641 standard; DNA: 353 BP.
 XX
 AC AAC56641;

DT 25-JAN-2001 (first entry)

XX Eucalyptus grandis transcription factor DNA sequence #512.

KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;

KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;

KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;

KW homeodomain zipper; LIM domain; AP2; ERBS; zinc finger domain;

KW type 2 Cys2His2; CCAT box element; MYB; ss.

OS Eucalyptus grandis.

PN WO200053724-A2.

PD 14-SEP-2000.

PF 09-MAR-2000; 2000WO-US06112.

XX 11-MAR-1999; 99US-026513.

PR 18-AUG-1999; 99US-0149485.

XX (GENE-) GENESTS RES & DEV CORP LTD.

PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

PI Wood M, McGrath A, Shenk MA, Glenn M;

PI WPI; 2000-579369/54.

XX Claim 1; Page 482; 747pp; English.

XX The present invention relates to novel plant transcription factors from

CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding

CC sequence for one such transcription factor. The transcription factor may

CC be used to produce a plant having modified gene expression such as a

CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or

CC mahogany species or to modify the activity of a polypeptide in a plant.

CC The transcription factors of the present invention are members from the

CC following families of regulatory proteins: bzip, bzip family of G-box
 CC binding factors, basic helix-loop-helix zipper, LIM domain, AP2
 CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
 CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.
 XX
 SQ Sequence 353 BP; 117 A; 59 C; 98 G; 78 T; 1 other;

Query Match 60.0%; Score 15; DB 21; Length 353;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 tcgtcaggtcaggt 21
 |||||
 Db 326 TCGTCAGGTCAGGT 312

RESULT 11
 AAC75563/c
 ID AAC75563 standard; cDNA; 586 BP.

AC AAC75563;

DT 08-FEB-2001 (first entry)

DE Human ORFX ORF1118 polynucleotide sequence SEQ ID NO:2235.

XX Human: open reading frame; ORFX; detection: cytosolic; hepatotropic;
 XX vulnerability: antiparkinsonian; neurotrophic; neuroprotective;
 XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiatic;
 XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 XX hypotensive; dermatological; immunosuppressive; antiinflammatory;
 XX antiviral; antibacterial; antifungal; antineumatic; antithyroid;
 XX antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
 XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 XX cholesterol ester storage; systemic lupus erythematosus; infection;
 XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 XX bone damage; cartilage damage; antiinflammatory disease; coagulation;
 XX thrombosis; contraceptive; ss.

OS Homo sapiens.

XX WO200058473-A2.

PD 05-OCT-2000.

PF 31-MAR-2000; 2000WO-US08621.

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

PA (CURA-) CURAGEN CORP.

PI Shinkets RA, Leach M;

DR WPI: 2000-602362/57.

DR P-PSDB: AAB41354.

XX Novel nucleic acids and peptides derived from open reading frame X,
 XX useful for treating e.g. cancers, proliferative disorders,
 XX neurodegenerative disorders and cardiovascular disease -

PS Claim 5; Page 1649; 5507bp; English.

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytosolic; hepatotropic; vulnerability;
 CC antiparkinsonian; neurotrophic; neuroprotective;

CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiatic; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antineumatic; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX

SQ Sequence 586 BP; 122 A; 185 C; 156 G; 123 T; 0 other;

Query Match 60.0%; Score 15; DB 21; Length 586;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 atctctgctcaggtc 17
 |||||
 Db 304 ATCTCTGCTCAGGTGC 290

RESULT 12
 AAX90421/c
 ID AAX90421 standard; cDNA; 4674 BP.

AC AAX90421;

DT 29-SEP-1999 (first entry)

DE Human ataxin-2 like gene.

XX Human: ataxin-2 like gene; ataxin-2 like protein; ataxia; deafness;
 XX cardiomyopathy; neurological disease; cancer; AIDS; diagnosis; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 63..3218

FT /tag= a
 /product= "ataxin-2 like protein"

PN WO9936527-A1.

PD 22-JUL-1999.

PF 19-JAN-1998; 98WO-CN00009.

PR 19-JAN-1998; 98WO-CN00009.

PA (UYHU-) UNIV HUNAN MEDICAL.

PI Deng H, Liu C, Wang D, Xia J;

DR WPI: 1999-458463/38.

DR P-PSDB: AAY29321.

XX Ataxin-2 like protein, and related polynucleotides, useful in
 XX treatment and diagnosis of ataxia,

PS Claim 4; Page 21-23; 34pp; English.

CC The present sequence represents a human ataxin-2 like gene. Ataxin-2
 CC like polynucleotides (1), vectors containing (1) and recombinant host
 CC cells are useful for recombinant production of Ataxin-2 like protein.

CC Both (I), ataxin-2 like protein and antibodies against ataxin-2 like
 CC protein are useful as research reagents, for screening assays and in
 CC diagnostic assays. Antagonists and agonists of ataxin-2 like protein
 CC can be used to inhibit or enhance, respectively, the activity of
 CC ataxin-2 like protein or expression of (I). Anti-ataxin-2 like protein
 CC antibodies and ataxin-2 like protein or its fragments can be used
 CC in vaccines. In particular, the proteins, antibodies, agonists and
 CC antagonists can be used for treating, e.g. ataxia, cardiomyopathy,
 CC deafness, neurological disease, cancer and AIDS, related to both an
 CC excess and insufficient amounts of ataxin-2 like.

Sequence 4674 BP: 971 A; 1566 C; 1244 G; 893 T; 0 other:

Query Match 60.0%; Score 15; DB 20; Length 4674;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 11 caggtcaggtcagc 25
 ||||||||||||
 Db 3741 CAGGTCCAGGTCCAGC 3727

RESULT 13
 AAH43851/C
 ID AAH43851 standard; cDNA: 6372 BP.

AC AAH43851;
 XX 04-SEP-2001 (first entry)
 XX

DE Human CLASP-7 encoding cDNA sequence SEQ ID NO:1.

XX Human: CLASP-7; cadherin-like asymmetry protein; immune response;
 KW neuroprotective; antidiabetic; immunosuppressive; antineumatic;
 KW antiarthritic; hypotensive; anti-HIV; cytostatic; immunostimulant;
 KW antianaemic; antiinflammatory; ophthalmological; nephroretrophic;
 KW antihypertoid; antiaesthetic; antiallergic; antibacterial; gene therapy;
 KW chromosome 19q13.2; autoimmune disease; multiple sclerosis; toxemia;
 KW juvenile diabetes; rheumatoid arthritis; pruritic urticarial papules;
 KW hypertension; Rh incompatibility; ss.

OS Homo sapiens.

XX Key Location/Qualifiers
 FH 13..6156
 FT /*tag= a
 FT /product= "CLASP-7"

XX WO200142295-A2.

XX 14-JUN-2001.

XX 13-DEC-2000; 2000WO-US34152.

XX 13-DEC-1999; 9905-0170453.

XX 14-JAN-2000; 2000US-0176195.

XX 14-FEB-2000; 2000US-0182296.

XX 11-APR-2000; 2000US-0196267.

XX 11-APR-2000; 2000US-0196460.

XX 11-APR-2000; 2000US-0196527.

XX 11-APR-2000; 2000US-0196528.

XX 11-APR-2000; 2000US-0547276.

XX 13-OCT-2000; 2000US-0240503.

XX 13-OCT-2000; 2000US-0240508.

XX 13-OCT-2000; 2000US-0240539.

XX 13-OCT-2000; 2000US-0240543.

(ARBO-) ARBOR VITA CORP.

Lu P, Garman JD, Candia AF;

WPI; 2001-381641/40.

DR P-PSDB; AAB9541.

XX Novel cadherin-like asymmetry protein-7 and polynucleotides encoding

PT the polypeptide, useful for treating autoimmune disease,

PT hypersensitivity, preventing transplant rejection by modulating immune

PT response

PS Claim 1; Fig 5; 151pp; English.

CC The present invention describes a human cadherin-like asymmetry protein
 CC (CLASP), designated CLASP-7. The CLASP-7 protein (I) and its encoding
 CC nucleotide sequence (II) have activities including: neuroprotective;
 CC antidiabetic; immunosuppressive; antineumatic; antiarthritic; anti-HIV;
 CC hypotensive; cytostatic; immunostimulant; antianaemic; antiinflammatory;
 CC ophthalmological; nephroretrophic; antihypertoid; antiaesthetic;
 CC antiallergic; and antibacterial. (II) and CLASP-7 antibodies (III) are
 CC useful for detecting the CLASP-7 polypeptide. (II) is useful for
 CC producing (I) by recombinant methods. (I) or its fragment are useful for
 CC inhibiting an immune response in a cell such as T cell or B cell. A
 CC pharmaceutical composition (C), comprising (I) or (II), can be useful
 CC for treating CLASP-7-mediated disease such as an autoimmune disease
 CC caused or exacerbated by increased activity of T helper cells. Autoimmune
 CC diseases which can be treated using (C) include multiple sclerosis,
 CC juvenile diabetes and rheumatoid arthritis. (I) is useful for treating
 CC toxemia or pregnancy induced hypertension, pruritic urticarial papules
 CC and Rh incompatibility. (I) is also useful as a diagnostic reagent for
 CC immune and other disorders, since diseases characterised by
 CC overproduction or depletion of lymphocytes in blood or other organs may
 CC be detected by monitoring the level of (I) or its mRNA. CLASP-7 has been
 CC mapped to the chromosomal location 19q13.2. The present sequence
 CC encodes the human CLASP-7 protein as given in the present invention.

XX Sequence 6372 BP: 1273 A; 2050 C; 1849 G; 1200 T; 0 other:

Query Match 60.0%; Score 15; DB 22; Length 6372;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 3 atctctcaggtcagc 17
 ||||||||||||
 Db 735 ATCTCTCAGGTCCAGTCC 721

RESULT 14

AAAC00048/C
 ID AAC00048 standard; cDNA: 327 BP.

XX AAC00048;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 46.

XX Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

XX gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EPI033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 9905-0122487.

XX (GENET) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

WPI; 2000-500381/45.

P-PSDB; AAG00042.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
PS Claim 1; SEQ ID 46; 71bp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.

CC Sequence 327 BP; 70 A; 102 C; 75 G; 69 T; 11 other;

XX

XX Query Match 56.0%; Score 14; DB 21; Length 327;
Best Local Similarity 100.0%; Pred. NO. 1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtcagtcagc 25
|||||
DB 144 AGCTGCAGCTCAGC 131

RESULT 15
AAH42390/C
ID AAH42390 standard; DNA; 363 BP.

XX AAH42390:
AC
XX
XX 01-OCT-2001 (first entry)
XX
XX
XX Nucleotide sequence of variable heavy chain fragment of clone G81.
DE
XX
XX Antibody; heavy chain; VH; amyloid protein; blood brain barrier;
KW endothelial cell; brain cell antigen; inflammation; adhesion molecule;
KW transferrin receptor; neurological disease; Alzheimer's disease;
XX prion disease; AIDS-related dementia; epilepsy; brain injury; ss.
XX
OS Homo sapiens.
XX
XX WO200144300-A2.
XX
XX 21-JUN-2001.
XX
XX 27-NOV-2000; 2000WO-CB04501.
XX
XX 13-DEC-1999; 99US-0170599.
XX
XX (CAMP-) CAMBRIDGE ANTIBODY TECHNOLOGY.
XX
XX Webster C, Osbourn J, Ward G, Miller K;
XX
XX WPI: 2001-398131/42.
XX
XX P-PSDB: AAG62946.
XX
XX Mixture or panel of antibodies for selecting specific binding members
PT that cross the blood brain barrier, for use in delivering different
PT molecules and treating neurological diseases -
XX
XX Disclosure; Page 98; 109pp; English.
XX
XX The present sequence encodes an antibody variable heavy chain (VH)

CC fragment. The fragment is used to produce a mixture or panel of 5
CC different specific binding members, each comprising an antibody VH
CC and/or VL variable domain and capable, when displayed on the surface
CC of filamentous bacteriophage particles or in the case of a specific
CC binding member comprising the D5 VH and/or VL variable domain when
CC bound to human serum amyloid protein, to pass through a mammalian
CC blood brain barrier (BBB). The panel is useful for the selection of
CC specific binding members with a desired property such as ability to
CC cross BBB, ability to bind endothelial cells or other brain cell antigen,
CC ability to bind areas of inflammation in the brain or BBB breakdown or
CC ability to bind intracellular adhesion molecules and to bind transferrin
CC receptor. The antibodies are useful in diagnosis, prophylaxis and
CC treatment of human or animal body, including neurological diseases, such
CC as Alzheimer's disease, prion disease, AIDS-related dementia, epilepsy
CC and traumatic brain injury and any diseases involving inflammation
CC occurring within the brain or central nervous system.

CC Sequence 363 BP; 84 A; 108 C; 96 G; 75 T; 0 other;

XX

XX Query Match 56.0%; Score 14; DB 22; Length 363;
Best Local Similarity 100.0%; Pred. NO. 1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtcagtcagc 25
|||||
DB 70 AGCTGCAGCTCAGC 57

RESULT 16
AAH34737/C
ID AAH34737 standard; cDNA; 370 BP.

XX AAH34737:
AC
XX
XX 03-SEP-2001 (first entry)
XX
XX
XX Human colon cancer antigen encoding cDNA SEQ ID NO:1819.
DE
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
XX
OS Homo sapiens.
XX
XX WO200122920-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26524.
XX
XX 29-SEP-1999; 99US-0157137.
XX
XX 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX WPI: 2001-235357/24.
XX
XX P-PSDB: AAG75332.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 1; Page 3364; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions

CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAH77789 represent sequences used in the exemplification of the
CC present invention.

CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX

S0 Sequence 370 BP; 112 A; 55 C; 81 G; 119 T; 3 other:

CC blood o(tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
XX dysplasia and hyperplasia.

50 Sequence 374 BP; 71 A; 116 C; 118 G; 68 T; 1 other;

Query Match 56.0%; Score 14; DB 22; Length 374;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 14; Conservative 0; Mismatches 0; Idels 0; Gaps 0;

Query Match	56.0%;	Score 14;	DB 22;	Length 374;
Best Local Similarity	100.0%;	Pred. No. 1e+02;		
Matches 14; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0

Query Match	56.0%;	Score 14;	DB 22;	Length 370;
Best Local Similarity	100.0%;	Pred. No. 1e+02;		
Matches 14;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	6	ctcgtcaggtgcag	19
Db	317	CTCGTCAGGTGCAC	304

QY	12	aggtgcagtcacg	25
Db	180	AGGTGCAGTCACG	167

RESULT	18
AA045597/c	
ID	AA045597 standard: DNA: 423 BP

RESULT	17
AAf64981/c	
ID	AAf64981 standard; cDNA; 374 BP

XX	04-DEC-1993	(first entry)
DT		
XX,		

AC	AAF64981;
XX	
DT	09-APR-2001 (first entry)

Variable heavy antibody chain; human subgroup 4 germ-line; ss

DE Novel human polynucleotide, SEQ ID NO: 737
xx

XX
CC
normal sapiens.

Human; cytostatic; gene therapy; colon cancer; prostate cancer
breast cancer; Lung cancer; cancer detection; ss.

```

ET      CDS      7.423
ET      /*tag= a

```

PN W02001.02568-A2.

XX 24-JUN-1993
PD

PD 11-JAN-2001.

XX 13-DEC-1991

PF 30-JUN-2000; 2000WO-US18374.

AA 13-DEC-1997: 91WQ-ATT00583

PR	02-JUL-1999;	99US-0142310.
PR	02-JUL-1999;	99US-0142311.

PA (DOWC) DOW CHEM AUSTRALIA
XX

PA (CHIR) CHIRON CORP.

P1 JOHNSON KS, MEZES PS, RICH
XX

Year	XX	XX
1990	100	100
1991	100	100
1992	100	100
1993	100	100
1994	100	100
1995	100	100
1996	100	100
1997	100	100
1998	100	100
1999	100	100
2000	100	100
2001	100	100
2002	100	100
2003	100	100
2004	100	100
2005	100	100
2006	100	100
2007	100	100
2008	100	100
2009	100	100
2010	100	100
2011	100	100
2012	100	100
2013	100	100
2014	100	100
2015	100	100
2016	100	100
2017	100	100
2018	100	100
2019	100	100
2020	100	100
2021	100	100
2022	100	100
2023	100	100
2024	100	100
2025	100	100
2026	100	100
2027	100	100
2028	100	100
2029	100	100
2030	100	100
2031	100	100
2032	100	100
2033	100	100
2034	100	100
2035	100	100
2036	100	100
2037	100	100
2038	100	100
2039	100	100
2040	100	100
2041	100	100
2042	100	100
2043	100	100
2044	100	100
2045	100	100
2046	100	100
2047	100	100
2048	100	100
2049	100	100
2050	100	100
2051	100	100
2052	100	100
2053	100	100
2054	100	100
2055	100	100
2056	100	100
2057	100	100
2058	100	100
2059	100	100
2060	100	100
2061	100	100
2062	100	100
2063	100	100
2064	100	100
2065	100	100
2066	100	100
2067	100	100
2068	100	100
2069	100	100
2070	100	100
2071	100	100
2072	100	100
2073	100	100
2074	100	100
2075	100	100
2076	100	100
2077	100	100
2078	100	100
2079	100	100
2080	100	100
2081	100	100
2082	100	100
2083	100	100
2084	100	100
2085	100	100
2086	100	100
2087	100	100
2088	100	100
2089	100	100
2090	100	100
2091	100	100
2092	100	100
2093	100	100
2094	100	100
2095	100	100
2096	100	100
2097	100	100
2098	100	100
2099	100	100
2100	100	100

DR WPL: 1993-2141/3/26.
DR P-PSDB: AAR38315.

PI Reinhard C, Randazzo F, Ke

aa
aa
PT

PI Kita D, Garcia V, Jones LW

PT subgroups 4 dermline gene. Its

DR WPI; 2001-091805/10.

F1 diagnosis of treatment of cancer
XX

Library of polynucleotides

disclosure, figure 4; 100pp; English
XX

PS Claim 9, Page 650, 1046bp; English.

XX Cell Line B17.2 expresses an antibody utilising a variable light
CC chain encoded by a gene derived from Hm μ 4 VL and a variable heavy
CC chain which makes a stable VL and VH combination.
XX

The present sequence is one of 335.

sequence 423 bp; 98 A; 120 C; 102 G; 9 T; 6 other,

genes cultivated with a cancerous state of a mammalian cell and can detect colon, prostate, breast and lung cancer. The library can be used to produce probes for detection of mRNA and to produce additional copies of the polynucleotides. The probes can be used for chromosome mapping

```

Query Match      56.0%; Score 14; DB 14; Length 423;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0

```

CC or antisense oligonucleotides can be generated. The polynucleotides and CC their gene products are used as genetic or biochemical markers (e.g. in

```

RESULT 19
AAZ39427/C
ID AAZ39427 standard; DNA: 423 BP.
XX
AC AAZ39427;
XX
XX 29-FEB-2000 (first entry)
XX
DE Antibody B17X2 VH nucleotide sequence.
XX
XX Tumor associated sialylated glycoprotein: TAG-72; cancer antigen;
XX carcinoma lesion; diagnostic; cancer; antibody; human; B17X2;
XX anti-mouse antibody hypersensitivity reaction; ss.
XX
XX Homo sapiens.
XX
XX US5976845-A.
XX
XX 02-NOV-1999.
XX
XX 07-JUN-1995; 95US-0487743.
XX
XX 16-JUN-1994; 94US-0261354.
XX
XX 19-APR-1990; 90US-0510697.
XX
XX 20-OCT-1992; 92US-0964536.
XX
XX (DOWC ) DOW CHEM CO.
XX
XX Johnson KS, Richard RA, Mezes PS;
XX
XX WPI: 1999-619651/53.
XX
XX P-PSDB; AAY57179.
XX
XX Production of humanized anti-TAG-72 antibodies, used for the detection,
XX in vivo imaging and treatment of cancers -
XX
XX Disclosure: Fig 4A-B: 85pp; English.
XX
XX The invention relates to producing humanized anti-tumor associated
XX sialylated glycoprotein (TAG-72) antibodies (anti-TAG). The antibodies
XX have binding specificity for the cancer antigen TAG-72. These antibodies
XX have variable regions with VL segments derived from human subgroup IV
XX germline gene and a VH segment (encoded by the VhalphaTag germline gene)
XX which is capable of combining with the VL to form a three dimensional
XX structure having the ability to bind TAG-72. They can also be used for the in
XX vivo detection of carcinoma lesions. They can also be used for in vitro
XX diagnostics. They can also be modified with therapeutic agents e.g. a
XX radionuclide, drug, biological response modifier, toxin or another
XX antibody for the treatment of cancers. The humanized anti-TAG-72
XX antibodies can reduce harmful anti-mouse antibody hypersensitivity
XX reactions.
XX
XX Sequence 423 BP; 98 A; 120 C; 102 G; 97 T; 6 other;
XX
Query Match 56.0%; Score 14; DB 20; Length 423;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 12 aggtgcagtcagc 25
DB 133 AGGTGCAGTCAGC 120

```

```

XX
XX Human B17X2 antibody VH segment DNA.
XX
XX Human; antibody; humanized; anti-tumor; sialylated glycoprotein antibody;
XX TAG-72; cytostatic; cancer antigen; detection; carcinoma lesion.
XX
XX diagnostic; treatment; ss.
XX
XX Homo sapiens.
XX
XX US5976531-A.
XX
XX 02-NOV-1999.
XX
XX 16-JUN-1994; 94US-0261354.
XX
XX 19-APR-1990; 90US-0510697.
XX
XX 20-OCT-1992; 92US-0964536.
XX
XX (DOWC ) DOW CHEM CO.
XX
XX Johnson KS, Mezes PS, Richard RA;
XX
XX WPI: 1999-632731/54.
XX
XX P-PSDB; AAY50688.
XX
XX New humanized anti-TAG-72 antibodies, used for the detection, in vivo
XX imaging and treatment of cancers -
XX
XX Disclosure: Figure 4A-B: 83pp; English.
XX
XX This invention describes novel humanized anti-tumor associated
XX sialylated glycoprotein antibodies (TAG-72) which have cytostatic
XX activity. The antibodies have binding specificity for the cancer antigen
XX TAG-72. They can be used for the in vivo detection of carcinoma lesions.
XX They can also be used for in vitro diagnostics. They can also be modified
XX with therapeutic agents e.g. a radionuclide, drug, biological response
XX modifier, toxin or another antibody for the treatment of cancers. The
XX humanized anti-TAG-72 antibodies can reduce harmful anti-mouse antibody
XX hypersensitivity reactions.
XX
XX Sequence 423 BP; 98 A; 120 C; 102 G; 97 T; 6 other;
XX
Query Match 56.0%; Score 14; DB 20; Length 423;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 12 aggtgcagtcagc 25
DB 133 AGGTGCAGTCAGC 120

```

```

RESULT 21
AAT80181/C
ID AAT80181 standard; DNA: 444 BP.
XX
XX AAT80181;
XX
XX 02-APR-1998 (first entry)
XX
XX Monoclonal antibody 105AD7 heavy chain variable region DNA sequence.
XX
XX Monoclonal antibody 105AD7; complementarily determining region;
XX CDR; human; immune response; treatment; cancer; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..444
XX FT /*tag= a
XX FT /transl_except= (pos: 157..159, aa: Gly)
XX FT /transl_except= (pos: 160..162, aa: Val)
XX FT /product= "105AD7 heavy chain variable region"

```

XX PN WO9732021-A1.
 XX XX 04-SEP-1997.
 PD PD
 XX PF 28-FEB-1997; 97WO-GE00591.
 XX PF 29-FEB-1996; 96GB-0004321.
 PR 28-FEB-1996; 96GB-0004177.
 XX XX
 PA (CANC-) CANCER RES. CAMPAIGN TECHNOLOGY LTD.
 XX PI Durrant LG, Robins RA, Spendlove I;
 XX XX WPI: 1997-448689/41.
 DR P-PSDB; AAM26239.
 XX XX
 PT Nucleic acid encoding peptide derived from antibody 105AD7 - used to
 XX induce an immune response for treatment and prevention of cancer
 PS Claim 1; Fig 1A; 32pp; English.

CC This is the DNA sequence of the monoclonal antibody 105AD7 heavy chain
 CC variable region. 105AD7 is a human monoclonal anti-idiotypic antibody
 CC that mimics T cell epitopes on the tumour associated antigen gp72 and
 CC has been used for therapeutic vaccination. A peptide derived from this
 CC antibody contains the 105AD7 heavy chain variable region and a 105AD7
 CC kappa chain. The complementarity determining regions (CDR) of this
 CC peptide and its fragments can stimulate immune responses. The peptide
 CC can be used for screening for its functional immune responses. The peptide,
 CC the peptides, and the fragments, equivalents and mimetics are used to
 CC stimulate an immune response to a tumour antigen, for the treatment or
 CC prevention of tumours. Transfected host cells are used to produce this
 CC peptide and the vectors can be used to generate the peptide in vivo for
 CC stimulation of an immune response. The peptide contains promiscuous
 CC helper epitopes which stimulate a response against tumour antigens other
 CC than gp72. These epitopes may help a cytotoxic T lymphocyte response to
 CC any co-injected antigen.

XX SO Sequence 444 BP; 100 A; 134 C; 107 G; 103 T; 0 other;

Query Match 56.0%; Score 14; DB 18; Length 444;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtgcaggtcagc 25
 |||||
 DB 121 AGGTGCAGGTCAAGC 108

RESULT 22
 AAZ20406/C
 ID AAZ20406 standard; cDNA: 447 BP.
 XX XX

AC AAZ20406;

DT 19-NOV-1999 (first entry)

DE IgG antibody 2.4.4 heavy chain coding sequence.

XX Antibody: CD147; Igm: ABX-CBL; activated T-cell killing; leukaemia;
 KW activated B-cell; monocyte; graft versus host disease; therapy; cancer;
 KW organ transplant rejection disease; lymphoma; pancreatic disease;
 KW autoimmune disease; inflammatory disease; arthritis; binding site; ss.
 XX OS Homo sapiens.

XX Key Location/Qualifiers
 FT CDS 1..447
 FT /*tag= a
 FT /*note= "no stop codon given"
 XX

PN WO945031-A2.
 XX XX 10-SEP-1999.
 PD PD
 XX XX
 XX PF 03-MAR-1999; 99WO-US04583.
 XX PF 03-MAR-1998; 98US-0034607.
 PR 03-FEB-1999; 99US-0244253.
 XX XX
 PA (ABCE-) ABGENIX INC.
 XX PI Davis CG, Blacher RM, Corvalan JR, Culwell AR, Green LL, Hales J;
 PI Havrilla N, Ivanov VE, Lipant JA, Liu Q, Weber RF, Yang X;
 XX XX WPI: 1999-540816/45.
 DR P-PSDB; AAY34305.
 XX XX
 PT New monoclonal antibody, used for treating e.g. graft versus host
 XX disease, cancers, autoimmune diseases and inflammatory diseases
 PS Disclosure; Fig 30; 245pp; English.

CC This sequence encodes the heavy chain of an antibody of the
 CC invention. The antibody is a monoclonal antibody (Mab) with an isotype
 CC that fixes complement and a variable region that binds to the epitope on
 CC CD147 bound by the Igm Mab ABX-CBL, providing that the antibody is not
 CC CBL1. The Mab can selectively kill activated T-cells, activated B-cells
 CC or resting or activated monocytes. The products and methods can be used
 CC for treating diseases involving activated T-cells or B-cells or
 CC monocytes, e.g. graft versus host disease (GVHD), organ transplant
 CC rejection diseases (e.g. renal transplant, ocular transplant), cancers
 CC (e.g. cancers of the blood (e.g. leukaemia's and lymphomas) and
 CC pancreatic), autoimmune diseases (e.g. lupus), and inflammatory diseases
 CC (e.g. arthritis).

XX SO Sequence 447 BP; 96 A; 150 C; 106 G; 95 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 447;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtgcaggtcagc 25
 |||||
 DB 34 AGGTGCAGGTCAAGC 21

RESULT 23
 AAQ78943/C
 ID AAQ78943 standard; DNA: 613 BP.
 XX XX

AC AAQ78943;

DT 07-AUG-1995 (first entry)

DE Human immunoglobulin variable heavy chain #5.

XX Primer: PCR; amplify; human; immunoglobulin; variable; heavy chain;
 KW cosmid; placenta; vector; pUBel; E.coli; mammalian; ds.
 XX

OS Homo sapiens.

XX Key Location/Qualifiers
 FT CDS 71..513
 FT /*tag= a
 FT /product= human immunoglobulin variable heavy chain
 FT intron 117..202
 FT /*tag= b
 FT misc_signal 514..516
 FT /*tag= c
 FT /*note= "miscellaneous signal, does not conform to
 FT terminator or splice site sequence"
 XX

```

PN      MO9426895-A.
XX      24-NOV-1994.
PD      10-MAY-1993; 93WO-JP00603.
XX      10-MAY-1993; 93WO-JP00603.
XX      10-MAY-1993; 93WO-JP00603.
XX      (NIBS ) JAPAN TOBACCO INC.
XX      Honjo T, Matsuda F;
XX      WPI: 1995-006791/01.
DR      P-PSDB; AAR66299.
XX      DNA fragment comprising human immunoglobulin Vh genes - for the
PT      production of human immunoglobulin in mammalian hosts
PS      Disclosure; Page 36-37; 130pp; Japanese.
XX      A series of genes (AA078939-79002) encoding human immunoglobulin
CC      variable heavy chains. The genes were isolated and cloned from a series
CC      of cosmid constructs: Y202; Y103; Y21; Y6; Y24; 3-31; M84; M118 and M131,
CC      by PCR amplification using primers AA078917-38. The genes are subdivided
CC      into 5 families of Vh genes. The fragments cover a region of 800 kb.
CC      The DNA fragments were isolated from high molecular weight DNA from
CC      human placenta. The DNA was partially digested with TaqI restriction
CC      enzyme. The fragments were separated by gel electrophoresis and 35-45 kb
CC      fractions were collected. The fragments were ligated with ClaI-digested
CC      cosmid vector pUB81. The ligation products were in vitro packed and
CC      infected into E.coli 490A. The fragments were then subcloned by colony
CC      hybridisation. The Vh genes and the DNA fragments encoding them are
CC      useful in producing human immunoglobulin in mammalian hosts.
XX      Sequence 613 BP; 139 A; 194 C; 139 G; 141 T; 0 other:
SO

```

```

Query Match          56.0%; Score 14; DB 16; Length 613;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      12 aggtcaggtcagc 25
        |||||
DB      283 AGCTGCAGCTCAGC 270

```

```

RESULT 24
AA078948/c
ID      AA078948 standard; DNA: 630 BP.
XX
XX      AA078948;
AC
XX
XX      01-AUG-1995 (first entry)
DT
XX
DE      Human immunoglobulin Vh gene #10.
XX
XX      Primer: PCR; amplify; human: immunoglobulin; variable; heavy chain;
KW      cosmid; placenta; vector; pUB81; E.coli; mammalian; ds.
XX
XX      Homo sapiens.
OS
XX
XX      Key
FH      Location/Qualifiers
FT      71..495
FT      /tag= a
FT      /product= human immunoglobulin variable heavy chain
FT      114..199
FT      /tag= b
FT      322..324
FT      /tag= c
FT      /transl_except= unused termination codon
FT      373..375
FT      /tag= d
FT      /transl_except= unused termination codon

```

```

FT      misc_signal      496..498
FT      /tag= e
FT      /note= "miscellaneous signal, does not conform to
FT      termination or splice site sequence"

```

```

XX      MO9426895-A.
XX      24-NOV-1994.
XX      10-MAY-1993; 93WO-JP00603.
XX      10-MAY-1993; 93WO-JP00603.
XX      10-MAY-1993; 93WO-JP00603.
XX      (NIBS ) JAPAN TOBACCO INC.
XX      Honjo T, Matsuda F;
XX      WPI: 1995-006791/01.
DR      P-PSDB; AAR66304.
XX      DNA fragment comprising human immunoglobulin Vh genes - for the
PT      production of human immunoglobulin in mammalian hosts
PS      Claim 19; Page 43-44; 130pp; Japanese.
XX
XX      A series of genes (AA078939-79002) encoding human immunoglobulin
CC      variable heavy chains. The genes were isolated and cloned from a series
CC      of cosmid constructs: Y202; Y103; Y21; Y6; Y24; 3-31; M84; M118 and M131,
CC      by PCR amplification using primers AA078917-38. The genes are subdivided
CC      into 5 families of Vh genes. The fragments cover a region of 800 kb.
CC      The DNA fragments were isolated from high molecular weight DNA from
CC      human placenta. The DNA was partially digested with TaqI restriction
CC      enzyme. The fragments were separated by gel electrophoresis and 35-45 kb
CC      fractions were collected. The fragments were ligated with ClaI-digested
CC      cosmid vector pUB81. The ligation products were in vitro packed and
CC      infected into E.coli 490A. The fragments were then subcloned by colony
CC      hybridisation. The Vh genes and the DNA fragments encoding them are
CC      useful in producing human immunoglobulin in mammalian hosts.
XX      Sequence 630 BP; 141 A; 179 C; 148 G; 162 T; 0 other:
SO

```

```

Query Match          56.0%; Score 14; DB 16; Length 630;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      12 aggtcaggtcagc 25
        |||||
DB      280 AGCTGCAGCTCAGC 267

```

```

RESULT 25
AA015905/c
ID      AA015905 standard; cDNA: 981 BP.
XX
XX      AA015905;
AC
XX
XX      12-JUN-2000 (first entry)
DT
XX
DE      Human protein clone HP10568 coding sequence.
XX
XX
XX      Human protein: hydrophobic domain; nutritional source; haematopoiesis;
KW      cytokine production; cell proliferation; cell differentiation;
KW      immune deficiency; infectious disease; autoimmune disorder; asthma;
KW      multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KW      allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
KW      nervous system disorder; Alzheimer's disease; Parkinson's disease;
KW      Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
KW      systemic cytokine damage; tissue differentiation; contraceptive; stroke;
KW      coagulation disorder; myocardial infarction; inflammatory condition;
KW      septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;
KW      nephritis; therapy; ss.
XX

```

OS Homo sapiens.
 XX
 PN WO200005367-A2.
 XX
 PD 03-FEB-2000.
 XX
 XX
 PF 22-JUL-1999; 99WO-JP03929.
 XX
 PR 24-JUL-1998; 98JP-0208820.
 PR 07-AUG-1998; 98JP-0224105.
 PR 25-AUG-1998; 98JP-0238116.
 PR 09-SEP-1998; 98JP-0254736.
 PR 29-SEP-1998; 98JP-0275505.
 XX
 PA (SAGA) SAGAMI CHEM RES CENT.
 PA (PROT-) PROTEGENE INC.
 XX
 PI Kato S, Kimura T;
 DR WPI: 2000-182694/16.
 DR P-PSDB: AAV94857.
 XX
 PT Novel human proteins having hydrophobic domains useful for treating
 PT osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,
 PT multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -
 XX
 PS Claim 3; Page 189; 351pp; English.
 XX
 XX This sequence encodes a human protein of the invention, which has
 CC hydrophobic domains. The DNA sequences can be used as a probe or as a
 CC genetic marker. The protein can also be used as a marker, and to identify
 CC potential genetic disorders. The DNA and protein can also be used as
 CC nutritional sources or supplements. The protein exhibits cytokine, cell
 CC proliferation, cell differentiation activities and induces production of
 CC other cytokines in certain cell populations. The protein also exhibits
 CC immune stimulating or immune suppressing activity. It can be used in the
 CC treatment of various immune deficiencies and disorders, and to treat
 CC infectious diseases caused by viral, bacterial, fungal or other
 CC infections. The protein is also used for treating autoimmune disorders
 CC such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid
 CC arthritis. It is also useful in the treatment of allergic reactions and
 CC conditions such as asthma, and in immune suppression after organ
 CC transplantation. The protein is useful in regulation of haematopoiesis
 CC and consequently in the treatment of myeloid or lymphoid cell
 CC deficiencies. It is also used in compositions for tissue growth or
 CC regeneration. The protein is also used in the treatment of osteoporosis
 CC or osteoarthritis and in the treatment of periodontal disease and other
 CC system disorders such as Alzheimer's disease, Parkinson's disease, and
 CC Huntington's disease. They are useful for protection or regeneration and
 CC treatment of lung or liver fibrosis, reperfusion injury in various
 CC tissues, and conditions resulting from systemic cytokine damage. They are
 CC also used for promoting or inhibiting tissue differentiation. They are
 CC also used as contraceptives since they exhibit activin or inhibin related
 CC activities and as a fertility inducing therapeutic. They are used for
 CC treating various coagulation disorders and in treatment and prevention of
 CC conditions resulting from coagulation activities e.g. myocardial
 CC infarction or stroke. They also acts as receptors, receptor ligands or
 CC inhibitors or agonists of receptor/ligand interactions. They are used to
 CC treat inflammatory conditions such as septic shock, sepsis, ischaemia
 CC reperfusion injury, arthritis, and nephritis. They can be used to
 CC prevent tumours.
 XX
 SQ Sequence 981 BP; 194 A; 299 C; 275 G; 213 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 981;
 Best Local Similarity 100.0%; Pred. No. 95;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtgcagctcagc 25
 |||||
 Db 139 AGGTGCAGCTCAGC 126

RESULT 26
 AAA15915/c
 ID AAA15915 standard; cDNA; 1100 BP.
 XX
 XX AAA15915;
 XX
 DT 12-JUN-2000 (First entry)
 XX
 DE Human protein clone HP10568 full length coding sequence.
 XX
 KW Human protein; hydrophobic domain; nutritional source; haematopoiesis;
 KW cytokine production; cell proliferation; cell differentiation;
 KW immune deficiency; infectious disease; autoimmune disorder; asthma;
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
 KW allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
 KW nervous system disorder; Alzheimer's disease; Parkinson's disease;
 KW Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
 KW systemic cytokine damage; tissue differentiation; contraceptive; stroke;
 KW coagulation disorder; myocardial infarction; inflammatory condition;
 KW septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;
 KW nephritis; therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200005367-A2.
 XX
 PD 03-FEB-2000.
 XX
 PF 22-JUL-1999; 99WO-JP03929.
 XX
 PR 24-JUL-1998; 98JP-0208820.
 PR 07-AUG-1998; 98JP-0224105.
 PR 25-AUG-1998; 98JP-0238116.
 PR 09-SEP-1998; 98JP-0254736.
 PR 29-SEP-1998; 98JP-0275505.
 XX
 PA (SAGA) SAGAMI CHEM RES CENT.
 PA (PROT-) PROTEGENE INC.
 XX
 PI Kato S, Kimura T;
 DR WPI: 2000-182694/16.
 DR P-PSDB: AAV94857.
 XX
 PT Novel human proteins having hydrophobic domains useful for treating
 PT osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,
 PT multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -
 XX
 PS Claim 4; Page 203-205; 351pp; English.
 XX
 XX This sequence encodes a human protein of the invention, which has
 CC hydrophobic domains. The DNA sequences can be used as a probe or as a
 CC genetic marker. The protein can also be used as a marker, and to identify
 CC potential genetic disorders. The DNA and protein can also be used as
 CC nutritional sources or supplements. The protein exhibits cytokine, cell
 CC proliferation, cell differentiation activities and induces production of
 CC other cytokines in certain cell populations. The protein also exhibits
 CC immune stimulating or immune suppressing activity. It can be used in the
 CC treatment of various immune deficiencies and disorders, and to treat
 CC infectious diseases caused by viral, bacterial, fungal or other
 CC infections. The protein is also used for treating autoimmune disorders
 CC such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid
 CC arthritis. It is also useful in the treatment of allergic reactions and
 CC conditions such as asthma, and in immune suppression after organ
 CC transplantation. The protein is useful in regulation of haematopoiesis
 CC and consequently in the treatment of myeloid or lymphoid cell
 CC deficiencies. It is also used in compositions for tissue growth or
 CC regeneration. The protein is also used in the treatment of osteoporosis
 CC or osteoarthritis and in the treatment of periodontal disease and other
 CC system disorders such as Alzheimer's disease, Parkinson's disease, and

CC Huntington's disease. They are useful for protection or regeneration and
CC treatment of lung or liver fibrosis, reperfusion injury in various
CC tissues, and conditions resulting from systemic cytokine damage. They are
CC also used for promoting or inhibiting tissue differentiation. They are
CC also used as contraceptives since they exhibit activin or inhibin related
CC activities and as a fertility inducing therapeutic. They are used for
CC treating various coagulation disorders and in treatment and prevention of
CC conditions resulting from coagulation activities e.g. myocardial
CC infarction or stroke. They also acts as receptors, receptor ligands or
CC inhibitors or agonists of receptor/ligand interactions. They are used to
CC treat inflammatory conditions such as septic shock, sepsis, ischaemia
CC reperfusion injury, arthritis, and nephritis. They can be used to
CC prevent tumours.

CC Sequence 1100 BP; 221 A; 333 C; 314 G; 232 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1100;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aggtcaggtcagc 25
|||||
DB 195 AGGTGACAGTCAGC 182

RESULT 27
AAA39064/C
ID AAA39064 standard; cDNA; 1130 BP.

AC AAA39064;

DT 30-AUG-2000 (first entry)

DE Human secreted protein gene 13 SEQ ID NO:23.

XX Human; secreted protein; cytostatic; anti-proliferative; vulnery;
XX immunosuppressive; antibacterial; diagnosis; immune system; chemotaxis;
XX hyperproliferative disorder; infectious disease; tissue regeneration;
XX screening; food additive; preservative; wound healing;
XX hyper-vascular disease; chromosome 11; ss.

OS Homo sapiens.

PN WO200017222-A1.

PN 30-MAR-2000.

PE 22-SEP-1999; 99WO-US22012.

XX 23-SEP-1998; 98US-0101546.

PR 02-OCT-1998; 98US-0102895.

PA (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Rosen CA, Duan RD, Shi Y, Lafleur DW, Young PE, Ni J;
PI Komatsoulis G, Endress GA, Soppet DR;

XX WPI: 2000-283538/24.

DR P-PSDB: AAB08903.

PT Human secreted proteins and coding sequences useful in diagnostic and
PT therapeutic methods for disorders such as immune system or
PT proliferative disorders, related to the proteins -

PS Claim 1; Page 329-330; 416pp; English.

XX The polynucleotide sequences given in AAA39052 to AAA39088 encode the
CC human secreted proteins given in AAB08891 to AAB08984. The human secreted
CC proteins can have activities based on the tissues and cells they are
CC expressed in. Examples of the activities are: cytostatic;
CC anti-proliferative; immunosuppressive; antibacterial; and vulnery. The
CC secreted proteins and their related polynucleotide sequences are useful

CC for diagnostic and therapeutic methods useful for diagnosing and treating
CC disorders related to the secreted proteins. The proteins, and
CC polynucleotide sequences may be useful for treating disorders of the
CC immune system, hyperproliferative disorders, infectious disease,
CC regeneration of tissues, for chemotaxis and for screening molecules that
CC bind to the proteins. The proteins or polynucleotide sequences may be
CC used as food additives or preservatives, to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, co-factors or other nutritional components. Agonists or
CC antagonists of the proteins may be used to prevent scar tissue growth
CC during wound healing, and hyper-vascular diseases. AAA39043 to AAA39051
CC and AAB08890 are sequences used in the exemplification of the present
CC invention.

XX Sequence 1130 BP; 246 A; 334 C; 317 G; 233 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1130;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aggtcaggtcagc 25
|||||
DB 201 AGGTGACAGTCAGC 188

RESULT 28
AAA39087/C
ID AAA39087 standard; cDNA; 1135 BP.

AC AAA39087;

DT 30-AUG-2000 (first entry)

DE Human secreted protein gene 13 SEQ ID NO:46.

XX Human; secreted protein; cytostatic; anti-proliferative; vulnery;
XX immunosuppressive; antibacterial; diagnosis; immune system; chemotaxis;
XX hyperproliferative disorder; infectious disease; tissue regeneration;
XX screening; food additive; preservative; wound healing;
XX hyper-vascular disease; chromosome 11; ss.

OS Homo sapiens.

PN WO200017222-A1.

PN 30-MAR-2000.

PE 22-SEP-1999; 99WO-US22012.

XX 23-SEP-1998; 98US-0101546.

PR 02-OCT-1998; 98US-0102895.

PA (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Rosen CA, Duan RD, Shi Y, Lafleur DW, Young PE, Ni J;
PI Komatsoulis G, Endress GA, Soppet DR;

XX WPI: 2000-283538/24.

DR P-PSDB: AAB08926.

PT Human secreted proteins and coding sequences useful in diagnostic and
PT therapeutic methods for disorders such as immune system or
PT proliferative disorders, related to the proteins -

PS Claim 1; Page 347-348; 416pp; English.

XX The polynucleotide sequences given in AAA39052 to AAA39088 encode the
CC human secreted proteins given in AAB08891 to AAB08984. The human secreted
CC proteins can have activities based on the tissues and cells they are
CC expressed in. Examples of the activities are: cytostatic;
CC anti-proliferative; immunosuppressive; antibacterial; and vulnery. The
CC secreted proteins and their related polynucleotide sequences are useful

CC for diagnostic and therapeutic methods useful for diagnosing and treating
 CC disorders related to the secreted proteins. The proteins, and
 CC polynucleotide sequences may be useful for treating disorders of the
 CC immune system, hyperproliferative disorders, infectious disease,
 CC regeneration of tissues, for chemotaxis and for screening molecules that
 CC bind to the proteins. The proteins or polynucleotide sequences may be
 CC used as food additives or preservatives, to increase or decrease storage
 CC capbilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, co-factors or other nutritional components. Agonists or
 CC antagonists of the proteins may be used to prevent scar tissue growth
 CC during wound healing, and hyper-vascular diseases. AAA39043 to AAA39051
 CC and AAB08890 are sequences used in the exemplification of the present
 CC invention.

XX
 XX
 S0 Sequence 1135 BP; 243 A; 333 C; 323 G; 234 T; 2 other;

Query Match 56.0%; Score 14; DB 21; Length 1135;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtcaggtcagc 25
 |||
 Db 188 AGGTGACGCTCAGC 175

RESULT 29
 AAA29322
 ID .AAA29322 standard; cDNA: 1160 BP.

XX AAA29322;

DT 26-SEP-2000 (first entry)

DE Glycine max isoflavone O-methyltransferase clone srl.pK0015.b4.

XX Isoflavone O-methyltransferase: soybean; degradation: daidzein;
 KW Flavonoid; flower colour; growth; pollination; irradiation; ss.

XX Glycine max.

XX Key Location/Qualifiers
 FT CDS 11..1087
 FT /tag=a
 FT /product= Isoflavone_O-methyltransferase

XX WO200037656-A2.

XX 29-JUN-2000.

XX 20-DEC-1999; 99WO-US30338.

XX 21-DEC-1998; 98US-0113190.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Famodu OO, McGonigle B, Odell JT, Fader GM, Falco SC;

XX WPI: 2000-442680/38.

XX P-PSDB: AAY96584.

XX New polynucleotide encoding flavonoid biosynthetic enzymes for creating
 XX transgenic plants and for immunological screening of cDNA libraries

XX Claim 2; Page 32; 39pp; English.

XX AAA29321-25 encode isoflavone O-methyltransferases isolated from various
 CC soybean (Glycine max) tissues. cDNA libraries seq, srl, srlc and sllc
 CC were prepared from soybean embryo (19 days after flowering), root, 8-day
 CC old root and seed (25 days after flowering).
 CC Isoflavone O-methyltransferase catalyses the first step in degradation
 CC of daidzein. Suppression of this enzyme will yield higher concentrations
 CC of this beneficial isoflavone in, e.g. soybean seed. Flavonoids serve as

CC co-pigments in flower colour, stimulate pollen tube growth, attract
 CC pollinators, act as feeding deterrents and protect against UV
 CC irradiation in fruits and seeds. The cDNA and proteins can be used to
 CC isolate homologues, for immunological screening and for positive
 CC selection methods.

XX
 XX
 S0 Sequence 1160 BP; 361 A; 197 C; 254 G; 348 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1160;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ctgcaggtcagc 19
 |||
 Db 222 ctgcaggtcagc 235

RESULT 30
 AAA91019/c
 ID AAA91019 standard; DNA: 1161 BP.

XX AAA91019;

DT 05-APR-2001 (first entry)

DE Human secreted protein PRO7154 coding sequence.

XX Secreted protein: human; PRO protein; neoplastic cell growth; tumour;

XX proliferation; leukaemia; lymphoid malignancy; inflammatory disorder;

XX angiogenic disorder; immunologic disorder; PRO7154; ds.

XX Homo sapiens.

XX Key Location/Qualifiers
 FT CDS 102..1085
 FT /tag=a
 FT /product= PRO7154

XX WO200075317-A2.

XX 14-DEC-2000.

XX 15-MAY-2000; 2000WO-US13358.

XX 09-JUN-1999; 99US-0138385.

XX 20-JUL-1999; 99US-0144790.

XX 03-AUG-1999; 99US-0146843.

XX 10-AUG-1999; 99US-0148188.

XX 17-AUG-1999; 99US-0149320.

XX 17-AUG-1999; 99US-0149327.

XX 17-AUG-1999; 99US-0149396.

XX 20-AUG-1999; 99US-0150114.

XX 31-AUG-1999; 99US-0151700.

XX 31-AUG-1999; 99US-0151734.

XX (GETH) GENENTECH INC.

XX Botstein DA, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;

XX WPI: 2001-071075/08.

XX P-PSDB: AAY97585.

XX Antibodies against PRO polypeptides, useful for diagnosing and treating
 CC tumours are associated with gene amplification, neoplastic cell growth
 CC and proliferation in mammals -
 CC Claim 50; Fig 11; 143pp; English.

XX This sequence encodes a human PRO protein of the invention. The PRO
 CC proteins are secreted proteins. Antagonists or antibodies of PRO
 CC polypeptides are useful for diagnosing and treating tumours are
 CC associated with gene amplification, neoplastic cell growth and

CC proliferation in mammals, and those conditions characterised by
CC overexpression and/or activation of the amplified genes. Such conditions
CC include benign or malignant tumours (e.g. renal, liver, kidney, bladder,
CC breast, gastric, ovarian, colorectal, prostate, pancreatic, lung, vulval,
CC thyroid, hepatic carcinomas, sarcomas, glioblastomas and various head and
CC neck tumours), leukaemias and lymphoid malignancies; neuronal, glial,
CC astrocytal, hypothalamic, and other glandular, macrophageal, epithelial,
CC stromal and blastocoele disorders; and inflammatory, angiogenic and
CC immunologic disorders. These may further be used to qualitatively or
CC quantitatively detect the expression of proteins encoded by the
CC amplified genes, and in tumour diagnostics or prognostics. The PRO
CC polypeptide or its antagonist may be used for the preparation of a
CC medicament in the treatment of a condition, which is responsive to the
CC PRO polypeptide, its antagonist or anti-PRO antibody.
CC
XX
SQ Sequence 1161 BP; 235 A; 356 C; 327 G; 242 T; 1 other;

Query Match 56.0%; Score 14; DB 22; Length 1161;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 aggtcagatcagc 25
|||||
Db 240 AGGTGACAGTCAGC 227

RESULT 31
AAZ98136/c
ID AAZ98136 standard; CDNA: 1173 BP.
XX
XX AAZ98136;
XX
DT 11-MAY-2000 (first entry)
DE Human signal peptide containing protein HSP-28 CDNA SEQ ID NO:162.
XX
XX Human: signal peptide-containing protein; HSP; diagnosis: cancer;
XX inflammation; cardiovascular disease; anticancer; anti-inflammatory;
XX antimicrobial; neuroprotective; cardiovascular; hepatocytic;
XX antiaesthetic; gene therapy; cell proliferation; neurological disorder;
XX reproductive disorder; developmental disorder; arteriosclerosis;
XX cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
XX asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
XX Parkinson's disease; Huntington's disease; ovulatory defect;
XX muscular dystrophy; ss.
XX
XX Homo sapiens.
XX
XX WO200000610-A2.
XX
XX 06-JAN-2000.
XX
XX 25-JUN-1999; 99WO-US14484.
XX
XX 26-JUN-1998; 98US-0090762.
XX PR 31-JUL-1998; 98US-0094983.
XX PR 01-OCT-1998; 98US-0102686.
XX PR 11-DEC-1998; 98US-0112129.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;
XX Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;
XX Bandman O;
XX
XX WPI: 2000-160673/14.
XX P-PSDB: AAY87251.
XX
XX New human signal peptide-containing proteins useful in treatment,
XX prevention and diagnosis of e.g. cancer, inflammation and
XX cardiovascular disease -

PS Claim 9: Page 268; 327pp; English.
XX
XX AAZ98109 to AAZ98242 encode AAY87251 to AAY87357 which represent the
XX human signal peptide-containing proteins HSP-1 to HSP-134. HSPs have
XX anticancer, anti-inflammatory, antimicrobial, neurotropic, hepatotropic,
XX neuroprotective, cardiovascular and antiaesthetic activities, and can
XX be used in gene therapy. HSPs can be used to treat or prevent disorders
XX associated with decreased activity or function of HSP. Antagonists of
XX HSP are used to treat or prevent disorders associated with increased
XX activity or function of HSP. Such diseases include cell proliferation
XX (including cancer), inflammation, cardiovascular, neurological,
XX reproductive or developmental disorders, (e.g. arteriosclerosis,
XX cirrhosis, psoriasis), acquired immune deficiency syndrome, anaemia,
XX asthma, Crohn's disease, microbial or other infections, congestive or
XX ischemic heart disease, Alzheimer's, Parkinson's or Huntington's
XX diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSP
XX nucleic acids can be used for the recombinant production of HSP, for
XX detecting HSP in standard hybridisation and amplification assays (for
XX diagnosis and monitoring), in gene therapy, as antisense,
XX triplex-forming or ribozyme therapeutics, for detecting related sequences
XX or genetic variations, and for chromosomal mapping. HSP are also used to
XX raise specific antibodies (Ab) and to screen for agonists and
XX antagonists (potential therapeutic agents). Ab are used to diagnose, or
XX monitor, HSP-related diseases (in usual immunoassays), as therapeutic
XX antagonists, in competitive drug screens, and for purification of HSP
XX from natural sources.
XX
XX
SQ Sequence 1173 BP; 236 A; 363 C; 330 G; 244 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1173;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 aggtcagatcagc 25
|||||
Db 256 AGGTGACAGTCAGC 243

RESULT 32
AAC98064/c
ID AAC98064 standard; CDNA: 1186 BP.
XX
XX AAC98064;
XX
XX 09-MAR-2001 (first entry)
XX
XX Human colon cancer antigen nucleotide sequence SEQ ID NO:74.
XX
XX
XX Human: colon cancer; colon cancer antigen; diagnosis: detection;
XX identification; cytostatic; cardiovascular; neuroprotective; vulnery;
XX immunomodulatory; muscular; gynaecological; gastrointestinal;
XX nephrotropic; anti-infective; antibacterial; gene therapy; wound;
XX neural disorder; immune system disorder; muscular disorder;
XX reproductive disorder; gastrointestinal disorder; renal disorder;
XX infectious disease; cardiovascular disorder; ss.
XX
XX Homo sapiens.
XX
XX WO2000055351-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05883.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI: 2000-587534/55.
XX P-PSDB: AAB53307.

XX Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer -
 XX
 PS Claim 1: Page 520-521; 2104pp: English.
 XX
 CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB51234 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 CC vulnereary, nephrotoxic, anti-infective and antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins
 CC may also be used to prevent diseases such as neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAB54007 represent sequences used in the exemplification of the present
 CC invention.
 CC
 XX Sequence 1186 BP; 244 A; 361 C; 338 G; 242 T; 1 other;
 SQ
 Query Match 56.0%; Score 14; DB 21; Length 1186;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 12 aggtgcaggtcagc 25
 |||||
 DB 255 AGGTGCAGGTCAAGC 242
 RESULT 33
 AAH35018/c
 ID AAH35018 standard; cDNA; 1186 BP.
 XX
 AC AAH35018;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen encoding cDNA SEQ ID NO:2100.
 XX
 KM Human: colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma; chromosome 11; ss.
 XX
 OS Homo sapiens.
 OS
 XX WO200122920-A2.
 PN
 PD 05-APR-2001.
 PD
 XX 28-SEP-2000; 2000WO-US26524.
 PF
 XX 29-SEP-1999; 99US-0157137.
 PR
 XX 03-NOV-1999; 99US-0163280.
 PR
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Ruben SM, Barash SC, Birse CE, Rosen CA;
 PI
 XX WPI; 2001-235357/24.
 DR
 DR P-PSDB; AAG75613.
 XX
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 XX
 PS Claim 1: Page 3586; 9803pp; English.
 PS
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon

CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patients own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated P,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 CC
 XX Sequence 1186 BP; 244 A; 362 C; 338 G; 242 T; 0 other;
 SQ
 Query Match 56.0%; Score 14; DB 22; Length 1186;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 12 aggtgcaggtcagc 25
 |||||
 DB 255 AGGTGCAGGTCAAGC 242
 RESULT 34
 AAT98726
 ID AAT98726 standard; DNA; 1191 BP.
 XX
 AC AAT98726;
 XX
 DT 09-NOV-1998 (first entry)
 XX
 DE DNA encoding a S. pneumoniae protein of unknown function.
 XX
 KM Streptococcus pneumoniae protein; genetic immunisation; antagonist;
 KW immunological response; inoculation; antibody production; inhibitor;
 KW T cell immune response; antimicrobial compound; bacterial adhesion;
 KW extracellular matrix protein; protein-mediated cell invasion; wound;
 XX
 XX pathogenesis; ss.
 XX
 OS Streptococcus pneumoniae.
 OS
 XX
 FH Key Location/Qualifiers
 FT 596..979
 FT CDS /*tag= a
 FT
 XX WO9743303-A1.
 PN
 PD 20-NOV-1997.
 PD
 XX 14-MAY-1997; 97WO-US07950.
 PF
 XX 14-MAY-1996; 96US-0017670.
 PR
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 PA
 XX Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
 PI
 XX Stodola RK;
 PI
 XX WPI; 1998-008793/01.
 DR
 DR P-PSDB; AAM38682.
 XX
 PT Novel Streptococcus pneumoniae proteins and related DNA - useful for
 PT diagnosing anti-microbial agents for treatment of bacterial
 PT infections

CC such a polypeptide. They can also be used to screen for compounds which
 CC interact with and inhibit or activate such a polypeptide. The
 CC polypeptides (or DNA encoding them, via gene therapy) are also useful
 CC for inducing an immunological response in a mammal. The antagonists are
 CC useful to inhibit such bacterial polypeptides. The polypeptides are
 CC particularly useful to identify antimicrobial compounds and antibiotics.
 CC They are also useful to determine their role in pathogenesis of
 CC infection, dysfunction and disease.

XX Sequence 1634 BP; 430 A; 339 C; 385 G; 480 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 1634;
 Best Local Similarity 100.0%; Pred. No. 91;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cgtcaggtgcaggt 21
 |||
 Db 954 cgtcaggtgcaggt 967

RESULT 37

AAV37360
 ID AAV37360 standard; DNA; 1634 BP.

AC AAV37360;

DT 13-OCT-1998 (first entry)

DE Streptococcus pneumoniae coding region.

KW coding region; ORF; open reading frame; antibacterial;
 KW infection; prevention; meningitis; ss.

OS Streptococcus pneumoniae.

XX Key Location/Qualifiers
 FT CDS 1040..1291
 FT /*tag= a
 FT /product= unknown

PN W09819689-A1.

PD 14-MAY-1998.

PE 27-OCT-1997; 97WO-US19226.

PR 01-NOV-1996; 96US-0029930.

PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.

PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
 PI Reid RH, Zarfos PN;

DR WPI; 1998-286586/25.
 DR P-PSDB: AAW60963.

XX New isolated nucleic acids from Streptococcus pneumoniae - useful,
 PT e.g. for identifying anti-bacterial(s) for treatment and prevention
 PT of meningitis

PS Claim 1; Page 60; 130pp; English.

XX The sequence is that of a coding region isolated from
 CC S. pneumoniae. Its encoded protein, or agonists of it,
 CC may be useful as an antibacterial for treatment or
 CC prevention of infection, specifically caused by S. pneumoniae
 CC (particularly meningitis) but possibly also Helicobacter
 CC pylori (ulcers and gastric cancer). It may be of particular
 CC use before insertion of an in-dwelling device or any other
 CC invasive procedure. The protein, or nucleic acid encoding
 CC it, can also be used in vaccines to induce a cellular

CC and/or humoral immune response, or to screen for other
 CC antibacterials. The DNA may also contain flanking sequences
 CC that are potential sources of control elements for bacterial
 CC gene expression. Detecting a sequence encoding the protein
 CC can be used diagnostically, e.g. to detect a mutation for
 CC serotyping or classifying infectious agents.

XX Sequence 1634 BP; 430 A; 339 C; 385 G; 480 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 1634;
 Best Local Similarity 100.0%; Pred. No. 91;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cgtcaggtgcaggt 21
 |||
 Db 954 cgtcaggtgcaggt 967

RESULT 38

AAV44597/C
 ID AAV44597 standard; DNA; 1758 BP.

AC AAV44597;

DT 24-NOV-1998 (first entry)

DE Mouse uncoupling protein-2 UCP2 gene promoter region.

XX uncoupling protein-2; UCP2 gene; mouse; respiration;

KW thermogenesis; obesity; hyperinsulinemia; glucose intolerance;
 KW diabetes; syndrome X; hypothermia; wasting; cachexia; anorexia;
 KW inflammation; fever; hyperthermia; gene therapy; diagnosis; ds.

XX Mus sp.

PN W09831396-A1.

PD 23-JUL-1998.

PE 22-APR-1997; 97WO-US06864.

PR 15-JAN-1997; 97US-0034960.

PA (NARE-) CENT NAT RECH SCI CENT RECH SUR ENDOCRINOL.
 PA (REGC) UNIT CALIFORNIA.
 PA (UYDU-) UNIT DUKE.

PI Boullaud F, Collins SA, Riquier D, Seldin MF;
 PI Surwit RS, Warden CH;

DR WPI; 1998-413823/35.

XX Method for treating disease associated with altered UCP-2 expression
 PT - by administering agent which enhances or inhibits UCP-2 activity,
 PT effectively to treat obesity, diabetes, fever, hyperthermia,
 PT cachexia etc.

PS Example II; Fig 8a; 98pp; English.

XX This is the nucleotide sequence of a fragment of the mouse
 CC uncoupling protein-2 (UCP2) gene, corresponding to the promoter
 CC region, obtained from genomic clone MMU2-L2. A DNA fragment from
 CC -934 to +8600 bp of the UCP2 gene has also been sequenced (see
 CC AAV44596). MMU2-L2 (deposited as 1-1868) was cloned from a mouse
 CC genomic library by screening with mouse UCP2 cDNA; the cDNA had
 CC been obtained from a mouse muscle cDNA library using rat UCP1 cDNA
 CC as probe. Human UCP2 cDNA and genomic sequences (see AAV4595,
 CC AAV4598-602 and AAV44629-38) are also provided, as well as a deduced
 CC amino acid sequence (see AAW69166) for human UCP2. The invention
 CC relates to a method for treating disease associated with altered
 CC UCP2 expression. An agent, which enhances UCP2 expression can be
 CC used to treat obesity, diabetes, syndrome X, hypothermia,

CC hyperinsulinaemia, or glucose intolerance. An inhibitor of UCP2
CC is used to treat wasting, anorexia, inflammation, cachexia, fever
CC or hyperthermia. The invention also relates to diagnostic and drug
CC screening methodologies.

XX Sequence 1758 BP; 421 A; 401 C; 438 G; 487 T; 11 other;

Query Match 56.0%; Score 14; DB 19; Length 1758;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 atctctgcacgtg 16
|||
Db 1072 ATCTCTGCACGTG 1059

RESULT 39

AA159259/c
ID AA159259 standard; cDNA: 1814 BP.

XX AA159259;

DT 22-OCT-2001 (first entry)

DE Human polynucleotide SEQ ID NO 1462.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokine; thrombolytic; drug screening; arthritis; inflammation;
XX Leukaemia; ss.

OS Homo sapiens.

XX WO200153312-A1.

PN 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US4263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSEQ-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX MPI: 2001-442253/47.

XX P-PSDB: AAM40103.

XX Novel nucleic acids and polypeptides, useful for creating disorders
XX such as central nervous system injuries -

XX Claim 1: SEQ ID NO 1462; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic/
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, hemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 1814 BP; 447 A; 531 C; 481 G; 355 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1814;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 agtgcacgtcagc 25
|||
Db 213 AGTGCACGTACAGC 200

RESULT 40

AA161045
ID AA161045 standard; cDNA: 1816 BP.

XX AA161045;

DT 22-OCT-2001 (first entry)

DE Human polynucleotide SEQ ID NO 5034.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokine; thrombolytic; drug screening; arthritis; inflammation;
XX Leukaemia; ss.

OS Homo sapiens.

XX WO200153312-A1.

PN 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US4263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSEQ-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX MPI: 2001-442253/47.

XX P-PSDB: AAM41889.

XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -

XX Claim 1: SEQ ID NO 5034; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic/
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide

XX Elmat P, Mor O, Skallier R, Feinstein E, Faerman A;
 XX WPI: 2000-053304/04.
 DR P-PSDB: AAY53669.
 XX
 PT Identification of stress induced genes for determining risk and
 PT preventing, treating or controlling osteoporosis
 XX
 PS Claim 25; Fig 9A-M: 308pp: English.
 XX
 CC The present sequence encodes protein 405, which was identified
 CC using the method of the invention after subjecting rat osteoblasts to
 CC mechanical stress. Expression of the 405 gene was found to be
 CC downregulated in response to mechanical stimulation and stress.
 CC The specification describes a method for the identification of genes
 CC responsive to a specific mechanical stress. The method comprises applying
 CC the mechanical stress to an organism (tissue or cells comprising bone
 CC cells), isolating the specific cellular fractions and extracting mRNA
 CC from them, and differentially analysing the mRNA in comparison with
 CC control samples. The method is used to identify genes whose expression
 CC is responsive to a specific stress. The identified genes are employed in
 CC determining risk associated with a physiological or disease state. The
 CC risk determination methods are used for testing a medicament for gene
 CC therapy. These medicaments, or genes identified by the method of the
 CC invention, are used for treating, preventing or controlling a
 CC physiological or disease state (especially osteoporosis or bone density
 CC or other factors causing or contributing to osteoporosis or its symptoms
 CC or other conditions involved in mechanical stress or its lack. The
 CC methods can also be used for advancing research or studies in bone
 CC development.
 S0 Sequence 3720 BP: 946 A; 991 C; 955 G; 828 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 3720;
 Best Local Similarity 100.0%; Pred. No. 85;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 gatccctgcagct 15
 |||||||
 Db 574 GATCCTGTCAGCT 561

RESULT 43
 AAS07382/C
 ID AAS07382 standard: cDNA: 4027 BP.
 XX
 AC AAS07382;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE Human DNA associated with CLASP-5 #2.
 XX
 KW Human: CLASP-5; cadherin-like asymmetry protein; immune gateway;
 KW immunogen; antibody; autoimmune disease; rheumatoid arthritis;
 KW multiple sclerosis; leukaemia; insulin dependent diabetes mellitus;
 KW acquired immunodeficiency syndrome; AIDS; SS.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..2970
 FT /ftag=a
 FT /product="Protein as displayed in AAU04026"
 FT /partial
 FT /note="No start codon"
 XX
 PN MO200142296-A2.
 XX
 PD 14-JUN-2001.
 XX
 PF 13-DEC-2000; 2000MO-US34163.

XX
 PR 13-DEC-1999; 99US-0170453.
 PR 14-JAN-2000; 2000US-0176195.
 PR 14-FEB-2000; 2000US-0182296.
 PR 11-APR-2000; 2000US-0196267.
 PR 11-APR-2000; 2000US-0196460.
 PR 11-APR-2000; 2000US-0196527.
 PR 11-APR-2000; 2000US-0196528.
 PR 11-APR-2000; 2000US-0547276.
 PR 13-OCT-2000; 2000US-0240503.
 PR 13-OCT-2000; 2000US-0240508.
 PR 13-OCT-2000; 2000US-0240543.
 PR 13-OCT-2000; 2000US-0240539.
 XX
 PA (ARBO-) ARBOR VITA CORP.
 XX
 PI Lu P, Garman JD, Candia AF;
 DR P-PSDB: AAU04026.
 XX
 DR WPI: 2001-367865/38.
 XX
 PT CLASP-5 polynucleotides, proteins and antibodies are used to prevent or
 PT treat a CLASP-5 mediated disease, such as an autoimmune disease e.g.
 PT Rheumatoid arthritis.
 XX
 PS Disclosure: Fig 9C: 188pp: English.
 XX
 CC The sequence encodes a protein presented as AAU04026, included in the
 CC specification which relates to a cadherin-like asymmetry protein,
 CC CLASP-5, which is a transmembrane protein of the immune system involved
 CC in the formation of the immune gateway. CLASP-5, polynucleotides encoding
 CC it and an anti-CLASP-5 antibody are used to prevent or treat a CLASP-5
 CC mediated disease, such as an autoimmune disease caused or exacerbated by
 CC increased activity of TH1 cells. These diseases may include Addison's
 CC disease, rheumatoid arthritis, dermatitis, Multiple Sclerosis, Systemic
 CC Lupus Erythematosus and Autoimmune Thyroiditis, inflammatory conditions
 CC (e.g. Ischaemia-reperfusion) and responses, Leukaemia, acquired
 CC immunodeficiency syndrome, AIDS, syphilis, typhoid, Burkitt's lymphoma,
 CC Grave's disease and insulin dependent diabetes mellitus. CLASP-5 is used
 CC for inhibiting an immune response in a cell (T cell or B cell) by
 CC interfering with the expression of a CLASP-5 gene in the cell, the
 CC ability of a CLASP-5 protein to bind to another cell or the ability of a
 CC CLASP-5 protein to bind to another protein. The antibody binding CLASP-5
 CC is used to inhibit an immune response is a subject. The polynucleotide is
 CC used to detect CLASP-5 expression in cells and for diagnosis of
 CC diseases and disorders associated with aberrant expression of CLASP-5.
 CC Note: The present sequence is included in the specification but is
 CC not mentioned anywhere in the specification.
 S0 Sequence 4027 BP: 1162 A; 898 C; 957 G; 1010 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 4027;
 Best Local Similarity 100.0%; Pred. No. 85;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 atccctgcagctg 16
 |||||||
 Db 1815 ATCCCTGTCAGCTG 1802

RESULT 44
 AAS07373/C
 ID AAS07373 standard: cDNA: 7215 BP.
 XX
 AC AAS07373;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE Human cDNA encoding CLASP-5.
 XX
 KW Human: CLASP-5; cadherin-like asymmetry protein; immune gateway;
 KW immunogen; antibody; autoimmune disease; rheumatoid arthritis;

KW multiple sclerosis; leukaemia; insulin dependent diabetes mellitus;
 KW acquired immunodeficiency syndrome; AIDS; ss.
 XX Homo sapiens.
 XX
 FH Key
 FT CDS
 FT
 FT Location/Qualifiers
 FT 112..6048
 FT /tag= a
 FT /product= "CLASP-5"
 FT 1498..1519
 FT /tag= b
 FT /note= "PCR primer HC59S1"
 FT 3087..3113
 FT /tag= c
 FT /note= "Antisense oligonucleotide AAS07384"
 FT 3087..3113
 FT /tag= d
 FT /note= "Antisense oligonucleotide AAS07384"
 FT 3232..3259
 FT /tag= e
 FT /note= "Antisense oligonucleotide AAS07384"
 FT complement (3642..3660)
 FT /tag= f
 FT /note= "PCR primer HC5AS10b"
 FT 4728..4756
 FT /tag= g
 FT /note= "Antisense oligonucleotide AAS07384"
 FT
 FT WO200142296-A2.
 XX
 XX 14-JUN-2001.
 XX
 XX
 XX 13-DEC-2000; 2000WO-US34163.
 XX
 XX 13-DEC-1999; 99US-0170453.
 XX 14-JAN-2000; 2000US-0176195.
 XX 14-FEB-2000; 2000US-0182296.
 XX 11-APR-2000; 2000US-0196267.
 XX 11-APR-2000; 2000US-0196460.
 XX 11-APR-2000; 2000US-0196527.
 XX 11-APR-2000; 2000US-0196528.
 XX 11-APR-2000; 2000US-0547276.
 XX 13-OCT-2000; 2000US-0240503.
 XX 13-OCT-2000; 2000US-0240508.
 XX 13-OCT-2000; 2000US-0240543.
 XX 13-OCT-2000; 2000US-0240539.
 XX
 PA (ARBO-) ARBOR VITA CORP.
 XX
 XX Lu P, Garman JD, Candia AF;
 PI
 XX WPI; 2001-367865/38.
 DR P-PSDB; AAU04024.
 XX
 XX
 PT CLASP-5 polynucleotides, proteins and antibodies are used to prevent or
 PT treat a CLASP-5 mediated disease, such as an autoimmune disease e.g.
 FT Rheumatoid arthritis -
 XX
 PS Claim 1; Fig 6; 188pp; English.
 XX
 CC The sequence encodes a cadherin-like asymmetry protein, CLASP-5, which
 CC is a transmembrane protein of the immune system involved in the formation
 CC of the immune gateway. CLASP-5, polynucleotides encoding it and an
 CC anti-CLASP-5 antibody are used to prevent or treat a CLASP-5 mediated
 CC disease, such as an autoimmune disease caused or exacerbated by
 CC increased activity of TH1 cells. These diseases may include Addison's
 CC disease, rheumatoid arthritis, dermatitis, Multiple Sclerosis, Systemic
 CC Lupus Erythematosus and Autoimmune Thyroiditis, inflammatory conditions
 CC (e.g. Ischaemia-reperfusion) and responses, Leukaemia, acquired
 CC immunodeficiency syndrome, AIDS, syphilis, typhoid, Burkitt's lymphoma,
 CC Grave's disease and insulin dependent diabetes mellitus. CLASP-5 is used
 CC for inhibiting an immune response in a cell (T cell or B cell) by
 CC interfering with the expression of a CLASP-5 gene in the cell, the

CC ability of a CLASP-5 protein to bind to another cell or the ability of a
 CC CLASP-5 protein to bind to another protein. The antibody binding CLASP-5
 CC is used to inhibit an immune response in a subject. The polynucleotide is
 CC used to detect CLASP-5 expression in cells and for diagnosis of
 CC diseases and disorders associated with aberrant expression of CLASP-5.
 XX
 SQ Sequence 7215 BP; 2005 A; 1734 C; 1721 G; 1755 T; 0 other;
 Query Match 56.0%; Score 14; DB 22; Length 7215;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 atctcgctcagctg 16
 Db 5004 ATCCTCGCTCAGCTG 4991
 RESULT 45
 ID AA011992
 AC AA011992 standard; DNA; 37856 BP.
 XX
 XX AA011992;
 XX 07-AUG-2000 (first entry)
 XX
 XX S. cellulosum DNA encoding polyketide and hereropolyketide enzymes.
 DE
 DE Polyketide biosynthesis; heteropolyketide biosynthesis; mutasynthesis;
 KW epothione; cytotoxic; immunosuppressant; antibiotic; antifungal;
 KW Plant-protection; ds.
 XX
 XX Sorangium cellulosum.
 OS
 XX
 FH Key
 FT CDS
 FT Location/Qualifiers
 FT /tag= a
 FT /product= "ORF1-tRNA synthetase"
 FT /note= "gtg start codon"
 FT complement (6374..7111)
 FT /tag= b
 FT /product= "ORF2-monoxygenase"
 FT complement (8433..9550)
 FT /tag= c
 FT /product= "ORF3-aminotransferase"
 FT /note= "AGT start codon given in the specification"
 FT 9855..11393
 FT /tag= d
 FT /product= "ORF4- tyrosine/DOPA-Decarboxylase"
 FT /note= "GNG start codon"
 FT 12212..13658
 FT /tag= e
 FT /product= "ORF5-3-oxoacyl-ACP-reductase"
 FT /note= "ACC start codon"
 FT 15374..19984
 FT /tag= f
 FT /product= "ORF6-polyketide synthase"
 FT 20003..27889
 FT /tag= g
 FT /product= "ORF7-peptide synthetase"
 FT 28251..29400
 FT /tag= h
 FT /product= "ORF8-transpeptidase"
 FT complement (30040..31720)
 FT /tag= i
 FT /product= "ORF9-regulation element"
 FT /note= "CGC stop codon"
 FT 31982..32932
 FT /tag= j
 FT /product= "ORF10-transcription regulator"
 FT 33128..33613
 FT /tag= k
 FT /product= "ORF11-regulation element"



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:55:38 ; Search time 143.1 Seconds
(without alignments)
39.566 Million cell updates/sec

Title: US-09-396-196f-9

Perfect score: 25

Sequence: 1 cgatccctcgcagtcgcagtcacg

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 9

Total number of hits satisfying chosen parameters: 12969

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : Issued Patents.NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCNUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfile1.seq.*

pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Match Length	ID	Description
	1	25	100.0	US-08-401-068-7	Sequence 7, Appli
	2	25	100.0	US-08-846-338-7	Sequence 7, Appli
	3	25	100.0	US-08-411-768B-1	Sequence 1, Appli
	4	25	100.0	US-08-411-768B-6	Sequence 6, Appli
	5	14	56.0	US-08-545-809A-5	Sequence 5, Appli
	6	14	56.0	US-08-545-809A-10	Sequence 10, Appli
	7	13	52.0	US-09-023-228B-96	Sequence 96, Appli
	8	13	52.0	US-08-483-636-59	Sequence 59, Appli
	9	13	52.0	US-08-483-632-59	Sequence 59, Appli
	10	13	52.0	US-08-483-632-63	Sequence 63, Appli
	11	13	52.0	US-08-483-632-63	Sequence 63, Appli
	12	13	52.0	US-08-488-376-13	Sequence 13, Appli
	13	13	52.0	US-08-488-376-13	Sequence 13, Appli
	14	13	52.0	US-08-634-223-13	Sequence 13, Appli
	15	13	52.0	US-08-634-223-13	Sequence 13, Appli
	16	13	52.0	US-08-634-224-13	Sequence 13, Appli
	17	13	52.0	US-08-634-224-13	Sequence 13, Appli
	18	13	52.0	US-08-634-400-13	Sequence 13, Appli
	19	13	52.0	US-08-634-400-13	Sequence 13, Appli
	20	13	52.0	US-08-635-878-13	Sequence 13, Appli
	21	13	52.0	US-08-635-878-13	Sequence 13, Appli
	22	13	52.0	US-08-770-057-13	Sequence 13, Appli
	23	13	52.0	US-08-770-057-13	Sequence 13, Appli
	24	13	52.0	US-09-335-697B-13	Sequence 13, Appli
	25	13	52.0	US-09-335-697B-13	Sequence 13, Appli
	26	13	52.0	US-08-345-321-9	Sequence 9, Appli
	27	13	52.0	US-08-545-809A-26	Sequence 26, Appli

28	13	52.0	750	4	US-09-183-959-18	Sequence 18, Appli
29	13	52.0	879	3	US-08-714-071-3	Sequence 3, Appli
30	13	52.0	1040	4	US-09-183-959-11	Sequence 11, Appli
31	13	52.0	1173	3	US-08-706-216-5	Sequence 5, Appli
32	13	52.0	1428	1	US-08-488-376-17	Sequence 17, Appli
33	13	52.0	1428	1	US-08-488-376-19	Sequence 19, Appli
34	13	52.0	1428	2	US-08-634-223-19	Sequence 19, Appli
35	13	52.0	1428	2	US-08-634-223-17	Sequence 17, Appli
36	13	52.0	1428	2	US-08-634-224-19	Sequence 19, Appli
37	13	52.0	1428	2	US-08-634-224-17	Sequence 17, Appli
38	13	52.0	1428	2	US-08-634-400-17	Sequence 17, Appli
39	13	52.0	1428	2	US-08-634-400-19	Sequence 19, Appli
40	13	52.0	1428	2	US-08-635-878-17	Sequence 17, Appli
41	13	52.0	1428	2	US-08-635-878-19	Sequence 19, Appli
42	13	52.0	1428	2	US-08-770-057-17	Sequence 17, Appli
43	13	52.0	1428	2	US-08-770-057-19	Sequence 19, Appli
44	13	52.0	1428	4	US-09-335-697B-17	Sequence 17, Appli
45	13	52.0	1428	4	US-09-335-697B-19	Sequence 19, Appli
46	13	52.0	2830	3	US-09-010-928B-1	Sequence 1, Appli
47	13	52.0	4533	3	US-08-726-214-5	Sequence 5, Appli
48	13	52.0	15239	1	US-08-390-878-17	Sequence 17, Appli
49	13	52.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
50	13	52.0	4411529	4	US-09-103-840A-1	Sequence 1, Appli
51	12	48.0	19	1	US-07-936-421-11	Sequence 11, Appli
52	12	48.0	20	3	US-08-545-809A-72	Sequence 72, Appli
53	12	48.0	28	2	US-08-353-372A-2	Sequence 2, Appli
54	12	48.0	50	2	US-08-850-049-122	Sequence 122, App
55	12	48.0	50	2	US-08-050-478-122	Sequence 122, App
56	12	48.0	50	4	US-09-414-117-122	Sequence 122, App
57	12	48.0	59	2	US-08-116-778B-24	Sequence 24, Appli
58	12	48.0	59	2	US-08-438-562-24	Sequence 24, Appli
59	12	48.0	59	2	US-08-483-528B-24	Sequence 24, Appli
60	12	48.0	61	4	US-08-673-799C-24	Sequence 24, Appli
61	12	48.0	61	4	US-08-687-421-385	Sequence 385, App
62	12	48.0	64	1	US-08-388-672A-8	Sequence 8, Appli
63	12	48.0	64	3	US-09-080-554-8	Sequence 8, Appli
64	12	48.0	66	1	US-08-207-996-5	Sequence 5, Appli
65	12	48.0	66	2	US-08-760-840A-5	Sequence 5, Appli
66	12	48.0	66	3	US-09-266-119-5	Sequence 5, Appli
67	12	48.0	70	1	US-08-290-592B-39	Sequence 39, Appli
68	12	48.0	70	5	PCT-US96-09448-39	Sequence 39, Appli
69	12	48.0	75	1	US-08-657-012-5	Sequence 5, Appli
70	12	48.0	75	4	US-09-013-872-5	Sequence 5, Appli
71	12	48.0	75	4	US-09-184-198-5	Sequence 5, Appli
72	12	48.0	76	4	US-09-025-769B-134	Sequence 134, App
73	12	48.0	93	4	US-08-556-978B-68	Sequence 68, Appli
74	12	48.0	93	4	US-08-467-420A-58	Sequence 58, Appli
75	12	48.0	138	1	US-08-470-110A-58	Sequence 58, Appli
76	12	48.0	138	1	US-08-667-769A-58	Sequence 58, Appli
77	12	48.0	138	2	US-08-940-371-58	Sequence 58, Appli
78	12	48.0	138	2	US-08-940-371-58	Sequence 58, Appli
79	12	48.0	138	5	PCT-US95-17082A-58	Sequence 58, Appli
80	12	48.0	162	1	US-08-275-916-1	Sequence 1, Appli
81	12	48.0	162	5	PCT-US93-03868-1	Sequence 1, Appli
82	12	48.0	198	2	US-08-414-657D-15	Sequence 15, Appli
83	12	48.0	198	2	US-08-414-657D-16	Sequence 16, Appli
84	12	48.0	303	4	US-08-556-978B-82	Sequence 82, Appli
85	12	48.0	357	1	US-08-459-310-11	Sequence 11, Appli
86	12	48.0	357	1	US-08-467-420A-61	Sequence 61, Appli
87	12	48.0	357	1	US-08-470-110A-61	Sequence 61, Appli
88	12	48.0	357	1	US-08-667-769A-61	Sequence 61, Appli
89	12	48.0	357	2	US-08-940-371-61	Sequence 61, Appli
90	12	48.0	357	2	US-08-483-636-67	Sequence 67, Appli
91	12	48.0	357	2	US-08-483-632-67	Sequence 67, Appli
92	12	48.0	357	5	PCT-US95-17082A-61	Sequence 61, Appli
93	12	48.0	359	4	US-09-025-769B-64	Sequence 64, Appli
94	12	48.0	407	2	US-08-353-372A-35	Sequence 35, Appli
95	12	48.0	429	1	US-08-463-128-3	Sequence 3, Appli
96	12	48.0	429	1	US-08-463-128-7	Sequence 7, Appli
97	12	48.0	429	1	US-08-463-298-3	Sequence 3, Appli
98	12	48.0	429	1	US-08-463-298-7	Sequence 7, Appli
99	12	48.0	429	1	US-08-436-339A-3	Sequence 3, Appli
100	12	48.0	429	2	US-08-436-339A-7	Sequence 7, Appli

ALIGNMENTS

```
RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 4.5e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctgcagtcagtcagtcagc 25
    |||
DB 105 CGATCCTGCTCAGTCAGTCAGTCAGC 129

RESULT 2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
```

```
ADDRESSEE: NO. 5869719artis Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Weigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 4.5e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctgcagtcagtcagtcagc 25
    |||
DB 105 CGATCCTGCTCAGTCAGTCAGTCAGC 129

RESULT 3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Johann Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: Of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect
; SOFTWARE: Version 5.1
; CURRENT APPLICATION DATA:
```

```

1  APPLICATION NUMBER: US/08/411,768B
2  FILING DATE: 31-March-95
3  CLASSIFICATION: 435
4  PRIOR APPLICATION DATA:
5  APPLICATION NUMBER: CH 3124/92
6  FILING DATE: 02-OCT-1992
7  PRIOR APPLICATION DATA:
8  APPLICATION NUMBER: CH 2134/93
9  FILING DATE: 15-JUL-1993
10 INFORMATION FOR SEQ ID NO: 1:
11 SEQUENCE CHARACTERISTICS:
12 LENGTH: 5872 base pairs
13 TYPE: nucleic acid
14 STRANDEDNESS: double
15 TOPOLOGY: linear
16 MOLECULE TYPE: DNA (genomic)
17 HYPOTHETICAL: NO
18 ANTI-SENSE: NO
19 ORIGINAL SOURCE:
20 ORGANISM: Escherichia coli
21 STRAIN: DSM498
22 IMMEDIATE SOURCE:
23 CLONE: pRO30A-15/9
24
25 FEATURE:
26 NAME/KEY: CDS
27 LOCATION: 117..1157
28 IDENTIFICATION METHOD: experimental
29 OTHER INFORMATION: /codon_start=117
30 OTHER INFORMATION: /product= "Biotin synthase"
31 OTHER INFORMATION: /evidence= EXPERIMENTAL
32 OTHER INFORMATION: /gene= "biob"
33 OTHER INFORMATION: /number= 1
34
35 FEATURE:
36 NAME/KEY: CDS
37 LOCATION: 2295..3050
38 OTHER INFORMATION: /codon_start=2295
39 OTHER INFORMATION: /function= "Involved in pimeoyl-CoA synthesis"
40 OTHER INFORMATION: /product= "protein"
41 OTHER INFORMATION: /gene= "bioc"
42 OTHER INFORMATION: /number= 3
43
44 FEATURE:
45 NAME/KEY: CDS
46 LOCATION: 3750..5039
47 IDENTIFICATION METHOD: experimental
48 OTHER INFORMATION: /codon_start=3750
49 OTHER INFORMATION: /EC_number= 2.6.1.62
50 OTHER INFORMATION: /product= "DAPA synthase"
51 OTHER INFORMATION: /evidence= EXPERIMENTAL
52 OTHER INFORMATION: /gene= "bioA"
53 OTHER INFORMATION: /number= 5
54 OTHER INFORMATION: /standard_name=
55 OTHER INFORMATION: "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
56 OTHER INFORMATION: aminotransf."
57
58 FEATURE:
59 NAME/KEY: CDS
60 LOCATION: 5098..5574
61 IDENTIFICATION METHOD: experimental
62 OTHER INFORMATION: /codon_start=5098
63 OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
64 OTHER INFORMATION: /product= "protein"
65 OTHER INFORMATION: /evidence= EXPERIMENTAL
66 OTHER INFORMATION: /gene= "orf1"
67 OTHER INFORMATION: /number= 6
68
69 FEATURE:
70 NAME/KEY: -10_signal
71 LOCATION: 45..49
72 IDENTIFICATION METHOD: experimental
73 OTHER INFORMATION: /evidence= EXPERIMENTAL
74 OTHER INFORMATION: /standard_name= "promoter plac"
75
76 FEATURE:
77 NAME/KEY: -35_signal
78 LOCATION: 23..28
79 OTHER INFORMATION: /standard_name= "promoter plac"

```

```

1 FEATURE:
2 NAME/KEY: RBS
3 LOCATION: 105..119
4 IDENTIFICATION METHOD: experimental
5 OTHER INFORMATION: /evidence= EXPERIMENTAL
6 OTHER INFORMATION: /standard_name= "b10b RBS no.9"
7 FEATURE:
8 NAME/KEY: RBS
9 LOCATION: 2284..2297
10 OTHER INFORMATION: /standard_name= "bioc RBS"
11 FEATURE:
12 NAME/KEY: RBS
13 LOCATION: 3742..3752
14 OTHER INFORMATION: /standard_name= "b10a RBS"
15 FEATURE:
16 NAME/KEY: RBS
17 LOCATION: 5088..5100
18 OTHER INFORMATION: /standard_name= "ORF1 RBS"
19 FEATURE:
20 NAME/KEY: terminator
21 LOCATION: 5583..5644
22 OTHER INFORMATION: /standard_name= "rho-independent"
23 OTHER INFORMATION: transcriptional terminator"
24 FEATURE:
25 NAME/KEY: stem_loop
26 LOCATION: 5583..5605
27 FEATURE:
28 NAME/KEY: promoter
29 LOCATION: 1..96
30 IDENTIFICATION METHOD: experimental
31 OTHER INFORMATION: /function= "promoter ptac"
32 OTHER INFORMATION: /evidence= EXPERIMENTAL
33 PUBLICATION INFORMATION:
34 DOCUMENT NUMBER: WO 87/01391 B1
35 FILING DATE: 26-AUG-1986
36 PUBLICATION DATE: 07-APR-1993
37 US-08-411-768B-1

```

```

Query Match Similarity      100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity      100.0%; Pred. No. 3.9e-05;
Matches      25; Conservative      0; Mismatches      0; Indels      0; Gaps      0.

Oy      1      cgatccctcgcagtcgaagtcagc      25
          |||||||
Db      221      CGATCCTCCTCAGTCGACGTCAGC      245

RESULT      4
: Sequence 6., Application US/0841768B
: Patent No. 6083712
: GENERAL INFORMATION:
: APPLICANT: Olwen Birch
: APPLICANT: Johann Brass
: APPLICANT: Martin Fuhrmann
: APPLICANT: Nicholas Shaw
: TITLE OF INVENTION: Biotechnological Method
: TITLE OF INVENTION: of Producing Bioclin
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
: STREET: 30 Rockefeller Plaza
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10112
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Wordperfect
: VERSION: 5.1

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pB030A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /EC_number=1154
OTHER INFORMATION: /EC_start=1154
OTHER INFORMATION: /EC_end=2308
OTHER INFORMATION: /product="KAPA synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="biof"
OTHER INFORMATION: /number=2
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /EC_number=3043
OTHER INFORMATION: /EC_start=3043
OTHER INFORMATION: /EC_end=3753
OTHER INFORMATION: /product="DTB synthase"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /gene="biod"
OTHER INFORMATION: /number=4
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name="biod RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name="biod RBS"
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6

Query Match
Best Local Similarity 100.0%; Score 25; DB 3; Length 5872;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 221 CGATCTCGTCAGGTGCGAGGTGAC 245

RESULT 5
US-08-545-809A-5/c
; Sequence 5, Application US/08545809A
; Patent No. 6096878
```

```

GENERAL INFORMATION:
APPLICANT: Honjo, Tasuku
APPLICANT: Matsuda, Fumihiko
TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
NUMBER OF SEQUENCES: 145
CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Richardson, F.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545,809A
FILING DATE: 27-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/00603
FILING DATE: 10-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 06501/004001
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 613 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL TYPE: human lymphoblast
CELL LINE: CGM1
US-08-545-809A-5

Query Match
Best Local Similarity 56.0%; Score 14; DB 3; Length 613;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 283 AGTGCAGGTGACG 270

RESULT 6
US-08-545-809A-10/c
; Sequence 10, Application US/08545809A
; Patent No. 6096878
GENERAL INFORMATION:
APPLICANT: Honjo, Tasuku
APPLICANT: Matsuda, Fumihiko
TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
NUMBER OF SEQUENCES: 145
CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Richardson, F.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
```


MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545.809A
FILING DATE: 27-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/00603
FILING DATE: 10-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 06501/004001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 630 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL TYPE: human lymphoblast
CELL LINE: CGM1
US-08-545-809A-10

Query Match 56.0%; Score 14; DB 3; Length 630;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 12 aggtcagatcagc 25
|||||
DB 280 AGGTGAGGTGAGC 267

RESULT 7
US-09-023-228B-96
Sequence 96, Application US/09023228B
Patent No. 6140490
GENERAL INFORMATION:
APPLICANT: BIESECKER, GREGORY
APPLICANT: GOLD, LARRY
TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS OF
TITLE OF INVENTION: COMPLEMENT SYSTEM PROTEINS
NUMBER OF SEQUENCES: 157
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Place #200
CITY: Denver
STATE: Colorado
COUNTRY: US
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023.228B
FILING DATE: 12-FEBRUARY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/011739
FILING DATE: 30 JAN 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/595,335
FILING DATE: 1 FEB 1996
ATTORNEY/AGENT INFORMATION:

NAME: Barry J. Swanson, Esq.
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX50/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 66 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA
FEATURE:
OTHER INFORMATION: All pyrimidines are 2'-F modified
US-09-023-228B-96

Query Match 52.0%; Score 13; DB 3; Length 66;
Best Local Similarity 76.9%; Pred. No. 98;
Matches 10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 cgaacctcgtcag 13
|||||
DB 41 CGAUCUCUCUCAG 53

RESULT 8
US-08-483-636-59/c
Sequence 59, Application US/08483636
Patent No. 5914110
GENERAL INFORMATION:
APPLICANT: Holmes, Stephen D.
APPLICANT: Gross, Mitchell S.
TITLE OF INVENTION: Recombinant IL4 Antibodies Useful in
TITLE OF INVENTION: Treatment of IL4 Mediated Disorders
NUMBER OF SEQUENCES: 75
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corp./Corporate
ADDRESSEE: Intellectual Property
STREET: P.O. Box 1539 / 0W2220
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483.636
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/117366
FILING DATE: 07-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/136783
FILING DATE: 14-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US/94/10308
FILING DATE: 07-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Sutton, Jeffrey A.
REGISTRATION NUMBER: 34,028
REFERENCE/DOCKET NUMBER: P50186-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5024
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:

LENGTH: 123 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-483-636-59

Query Match 52.0%; Score 13; DB 2; Length 123;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 aggtgcaggtcag 24
|||||
Db 70 AGGTGCAGGTGACG 58

RESULT 9

US-08-483-632-59/c
Sequence 59, Application US/08483632
Patent No. 5928904
GENERAL INFORMATION:
APPLICANT: Holmes, Stephen D.
APPLICANT: Gross, Mitchell S.
APPLICANT: Sylvester, Daniel R.
TITLE OF INVENTION: Recombinant IL4 Antibodies Useful in
TREATMENT OF ALLERGIC DISORDERS
NUMBER OF INVENTIONS: 75
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SmithKline Beecham Corp./Corporate
STREET: P.O. Box 1539 / UW2220
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,632
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/117366
FILING DATE: 07-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/136783
FILING DATE: 14-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US/94/10308
FILING DATE: 07-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Sutton, Jeffrey A.
REGISTRATION NUMBER: 34,028
REFERENCE/DOCKET NUMBER: P50186-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5024
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 123 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-483-632-59

Query Match 52.0%; Score 13; DB 2; Length 123;
Best Local Similarity 100.0%; Pred. No. 93;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 12 aggtgcaggtcag 24
|||||
Db 70 AGGTGCAGGTGACG 58

RESULT 10

US-08-483-636-63/c
Sequence 63, Application US/08483636
Patent No. 5914110
GENERAL INFORMATION:
APPLICANT: Holmes, Stephen D.
APPLICANT: Gross, Mitchell S.
APPLICANT: Sylvester, Daniel R.
TITLE OF INVENTION: Recombinant IL4 Antibodies Useful in
TREATMENT OF ALLERGIC DISORDERS
NUMBER OF INVENTIONS: 75
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SmithKline Beecham Corp./Corporate
STREET: P.O. Box 1539 / UW2220
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,636
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/117366
FILING DATE: 07-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/136783
FILING DATE: 14-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US/94/10308
FILING DATE: 07-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Sutton, Jeffrey A.
REGISTRATION NUMBER: 34,028
REFERENCE/DOCKET NUMBER: P50186-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5024
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 363 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-483-636-63

Query Match 52.0%; Score 13; DB 2; Length 363;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 aggtgcaggtcag 24
|||||
Db 70 AGGTGCAGGTGACG 58

RESULT 11

US-08-483-632-63/c
Sequence 63, Application US/08483632

Patent No. 5928904
GENERAL INFORMATION:
APPLICANT: Holmes, Stephen D.
APPLICANT: Gross, Mitchell S.
APPLICANT: Sylvester, Daniel R.
TITLE OF INVENTION: Recombinant IL4 Antibodies Useful in
TITLE OF INVENTION: Treatment of IL4 Mediated Disorders
NUMBER OF SEQUENCES: 75
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corp./Corporate
ADDRESS: Intellectual Property
STREET: P.O. Box 1539 / UM2220
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,632
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/117366
FILING DATE: 07-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/136783
FILING DATE: 14-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US/94/10308
FILING DATE: 07-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Sutton, Jeffrey A.
REGISTRATION NUMBER: 34,028
REFERENCE/DOCKET NUMBER: P50186-3
TELEPHONE: (215) 270-5024
TELEFAX: (215) 270-5024
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 363 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-483-632-63

Query Match 52.0%; Score 13; DB 2; Length 363;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aggtgcaggtcag 24
|||||
DB 70 AGCTGCAGGTCA 58

RESULT 12
US-08-488-376-13/c
Sequence 13, Application US/08488376
Patent No. 5811524
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND

TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,376
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-488-376-13

Query Match 52.0%; Score 13; DB 1; Length 378;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 ggtcaggtcag 25
|||||
DB 69 GTGTCAGGTCA 57

RESULT 13
US-08-488-376-15/c
Sequence 15, Application US/08488376
Patent No. 5811524
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,376
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-488-376-15

Query Match 52.0%; Score 13; DB 1; Length 378;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 agtgcaggtcagc 24
|||||
DB 70 AGTGCAGGTGACG 58

RESULT 14
US-08-634-223-13/c
Sequence 13, Application US/08634223
Patent No. 5840298
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,223
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-634-223-13

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ggtgcaggtcagc 25
|||||
DB 69 GGTGCAGGTGACG 57

RESULT 15
US-08-634-223-15/c
Sequence 15, Application US/08634223
Patent No. 5840298
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,223
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:

NAME/KEY: CDS
LOCATION: 1..378
US-08-634-223-15

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 agtgcaggtcag 24
|||||
DB 70 AGTGCAGGTCAG 58

RESULT 16
US-08-634-224-13/c
Sequence 13, Application US/08634224
Patent No. 5866125
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,224
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-6620
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-634-224-13

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 13 ggtgcaggtcag 25

|||||
DB 69 GGTGCAGGTCAG 57

RESULT 17
US-08-634-224-15/c
Sequence 15, Application US/08634224
Patent No. 5866125
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,224
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-634-224-15

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 agtgcaggtcag 24
|||||
DB 70 AGTGCAGGTCAG 58

RESULT 18
US-08-634-400-13/c
Sequence 13, Application US/08634400
Patent No. 5939068
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter

APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,400
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-634-400-13

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ggtagcagtcagc 25
|||||

Db 69 ggtagcagtcagc 57

RESULT 19
US-08-634-400-15/c
Sequence 15 Application US/08634400
Patent No. 5938068
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:

ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,400
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-634-400-15

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 agtgagtcag 24
|||||

Db 70 agtgagtcagtcag 58

RESULT 20
US-08-635-878-13/c
Sequence 13 Application US/08635878
Patent No. 5955364
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/635,878
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-635-878-13

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ggtgcaggtcagc 25
|||||
DB 69 ggtgcaggtcagc 57

RESULT 21
US-08-635-878-15/C
Sequence 15, Application US/08635878
Patent No. 5955364
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/635,878
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.

REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-635-878-15

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtgcaggtcagc 24
|||||
DB 70 aggtgcaggtcagc 58

RESULT 22
US-08-770-057-13/C
Sequence 13, Application US/08770057
Patent No. 5958765
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,057
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-770-057-13

Query Match
Best Local Similarity 52.0%; Score 13; DB 2; Length 378;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 ggtgcagtcacg 25
|||||
DB 69 GGTGCAGTCACG 57

RESULT 23

US-08-770-057-15/c
Sequence 15, Application US/08770057
Patent No. 5958765

GENERAL INFORMATION:

APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
NUMBER OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,057
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/488,376

FILING DATE: 07-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Teskin, Robin L.

REGISTRATION NUMBER: 35,030

REFERENCE/DOCKET NUMBER: 012712-150

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 378 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS
LOCATION: 1..378
US-08-770-057-15

Query Match
Best Local Similarity 52.0%; Score 13; DB 2; Length 378;
Matches 13; Conservative 100.0%; Pred. No. 85;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aggtgcagtcacg 24
|||||
DB 70 AGGTGCAGTCACG 58

RESULT 24

US-09-335-697B-13/c

Sequence 13, Application US/09335697B

Patent No. 6200804

GENERAL INFORMATION:

APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/335,697B
FILING DATE: 06-JUL-2000
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/770,057
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 378 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS
LOCATION: 1..378
US-09-335-697B-13

Query Match
Best Local Similarity 52.0%; Score 13; DB 4; Length 378;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 ggtgcagtcacg 25
|||||
DB 69 GGTGCAGTCACG 57

RESULT 25
US-09-335-697B-15/c

Sequence 15, Application US/09335697B
Patent No. 6200804
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
CHAMAT, Soulaïma Salim
PAN, Li-Zhen
WALSH, Edward E.
HEARD, Cheryl Janne
NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/335,697B
FILING DATE: 06-Jul-2000
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/770,057
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Rodin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-09-335-697B-15

Query Match 52.0%; Score 13; DB 4; Length 378;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 aggtgcaggtcag 24
|||||
Db 70 AGGTGCAGGTGAG 58

RESULT 26
US-08-345-321-9/c
Sequence 9, Application US/08345321
Patent No. 5914109
GENERAL INFORMATION:
APPLICANT: ZOLLA-PAZNER, Susan
GORNY, Mirosław K.
TITLE OF INVENTION: HETEROHYBRIDOMAS PRODUCING HUMAN
MONOCLONAL ANTIBODIES TO HIV-1
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:

ADDRESSEE: Browdy and Nelmark
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/345,321
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/872,675
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: ZOLLA-PAZNERIB
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..423
US-08-345-321-9

Query Match 52.0%; Score 13; DB 2; Length 423;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 aggtgcaggtcag 24
|||||
Db 127 AGGTGCAGGTGAG 115

RESULT 27
US-08-545-809A-26/c
Sequence 26, Application US/08545809A
Patent No. 6096878
GENERAL INFORMATION:
APPLICANT: Honjo, Tasuku
MATSUDA, Fumihiko
TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
NUMBER OF SEQUENCES: 145
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSO for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545,809A
FILING DATE: 27-MAR-1996

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/00603
FILING DATE: 10-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 06501/004001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 546 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL TYPE: human lymphoblast
CELL LINE: CGM1
US-08-545-809A-26

Query Match 52.0%; Score 13; DB 3; Length 546;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ggtgcaggtcagc 25
|||||
DB 232 GGTGCAGGTCAAC 220

RESULT 28
US-09-183-959-18
Sequence 18, Application US/09183959
Patent No. 6303332
GENERAL INFORMATION:
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Miao, Guo-Hua
APPLICANT: Rafalski, J. Antoni
TITLE OF INVENTION: CORN CDNA ENCODING SOUTHERN LEAF BLIGHT RESISTANCE
FILE REFERENCE: BB-1125
CURRENT APPLICATION NUMBER: US/09/183,959
EARLIER FILING DATE: 1998-11-02
EARLIER APPLICATION NUMBER: 60/064,493
NUMBER OF SEQ ID NOS: 20
SOFTWARE: Microsoft Windows 95
SEQ ID NO 18
LENGTH: 750
TYPE: DNA
ORGANISM: Zea mays
US-09-183-959-18

Query Match 52.0%; Score 13; DB 4; Length 750;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ggtgcaggtcagc 25
|||||
DB 687 ggtgcaggtcagc 699

RESULT 29
US-08-714-071-3/c
Sequence 3, Application US/08714071
Patent No. 6136584
GENERAL INFORMATION:
APPLICANT: Tsutsumi, FUJIWARA

APPLICANT: Shiro, OKUNO
APPLICANT: Hisanobu, HIRANO
APPLICANT: Sadahito, SHIN
TITLE OF INVENTION: FK506 BINDING PROTEIN GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,071
FILING DATE:
CLASSIFICATION: 435
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 879 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Human fetal brain cdna library
IMMEDIATE SOURCE:
CLONE: OTK4(6-1)
FEATURE:
NAME/KEY: CDS
LOCATION: 70..393
US-08-714-071-3

Query Match 52.0%; Score 13; DB 3; Length 879;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ggtgcaggtcagc 25
|||||
DB 303 GGTGCAGGTCAAC 291

RESULT 30
US-09-183-959-11
Sequence 11, Application US/09183959
Patent No. 6303332
GENERAL INFORMATION:
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Miao, Guo-Hua
APPLICANT: Rafalski, J. Antoni
TITLE OF INVENTION: CORN CDNA ENCODING SOUTHERN LEAF BLIGHT RESISTANCE
FILE REFERENCE: BB-1125
CURRENT APPLICATION NUMBER: US/09/183,959
EARLIER FILING DATE: 1998-11-02
EARLIER APPLICATION NUMBER: 60/064,493
NUMBER OF SEQ ID NOS: 20
SOFTWARE: Microsoft Windows 95
SEQ ID NO 11
LENGTH: 1040
TYPE: DNA

ORGANISM: Zee mays
US-09-183-959-11

Query Match 52.0%; Score 13; DB 4; Length 1040;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 ggtcaggtcagc 25
|||||
Db 426 ggtcaggtcagc 438

RESULT 31
US-08-706-216-5/c
; Sequence 5, Application US/08706216
; Patent No. 6140098
; GENERAL INFORMATION:
; APPLICANT: Balasubramanian, Sri Ram
; APPLICANT: Ford, John
; APPLICANT: Gorman, Daniel M.
; APPLICANT: Zurawski, Gerard
; TITLE OF INVENTION: MAMMALIAN PROTEASES; RELATED REAGENTS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/706,216
; FILING DATE: 30-AUG-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-852-9196
; TELEFAX: 415-496-1200
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1173 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1173
; US-08-706-216-5

Query Match 52.0%; Score 13; DB 3; Length 1173;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tcaggtcagtc 22
|||||
Db 418 TCAGTGTGAGTC 406

RESULT 32
US-08-488-376-17/c
; Sequence 17, Application US/08488376
; Patent No. 5811524

GENERAL INFORMATION:
; APPLICANT: BRAMS, Peter
; APPLICANT: CHAMAT, Soulaïma Salim
; APPLICANT: PAN, Li-Zhen
; APPLICANT: WALSH, Edward E.
; APPLICANT: HEARD, Cheryl Janne
; APPLICANT: NEWMAN, Roland Anthony
; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
; TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,376
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 012712-150
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1428 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1428
; US-08-488-376-17

Query Match 52.0%; Score 13; DB 1; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 ggtcaggtcagc 25
|||||
Db 126 GGTCTAGTCTAGC 114

RESULT 33
US-08-488-376-19/c
; Sequence 19, Application US/08488376
; Patent No. 5811524
; GENERAL INFORMATION:
; APPLICANT: BRAMS, Peter
; APPLICANT: CHAMAT, Soulaïma Salim
; APPLICANT: PAN, Li-Zhen
; APPLICANT: WALSH, Edward E.
; APPLICANT: HEARD, Cheryl Janne
; APPLICANT: NEWMAN, Roland Anthony
; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
; TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis

STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,376
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-488-376-19

Query Match 52.0%; Score 13; DB 1; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtcaggtcag 24
|||||

DB 127 AGTCTCAGCTCAG 115

RESULT 34
US-08-634-223-17/c
Sequence 17, Application US/08634223
Patent No. 5840298
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,223
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-634-223-17

Query Match 52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ggtcaggtcagc 25
|||||

DB 126 GGTCTCAGCTCAGC 114

RESULT 35
US-08-634-223-19/c
Sequence 19, Application US/08634223
Patent No. 5840298
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,223
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-634-223-19

Query Match
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtcaggtcag 24
|||||

DB 127 AGTGCAGGTCTAG 115

RESULT 36
US-08-634-224-17/C
Sequence 17, Application US/08634224
Patent No. 5866125
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,224
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS

LOCATION: 1..1428
US-08-634-224-17

Query Match
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ggtcaggtcagc 25
|||||

DB 126 GGTGCAGGTCTAGC 114

RESULT 37
US-08-634-224-19/C
Sequence 19, Application US/08634224
Patent No. 5866125
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,224
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-634-224-19

Query Match
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aggtcaggtcag 24
|||||

DB 127 AGTGCAGCTCAG 115

RESULT 38
US-08-634-400-17/c

; Sequence 17, Application US/08634400

; Patent No. 5939068

; GENERAL INFORMATION:

; APPLICANT: BRAMS, Peter

; APPLICANT: CHAMAT, Soulaïma Salim

; APPLICANT: PAN, Li-Zhen

; APPLICANT: WALSH, Edward E.

; APPLICANT: HEARD, Cheryl Janne

; APPLICANT: NEWMAN, Roland Anthony

; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN

; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND

; NUMBER OF SEQUENCES: 19 METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Burns, Doane, Swecker & Mathis

; STREET: P.O. Box 1404

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 2213-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentln Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/634,400

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/488,376

; FILING DATE: 07-JUN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Teskin, Robin L.

; REGISTRATION NUMBER: 35,030

; REFERENCE/DOCKET NUMBER: 012712-150

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 836-6620

; TELEFAX: (703) 836-2021

; INFORMATION FOR SEQ ID NO: 17:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1428 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..1428

; US-08-634-400-17

Query Match 52.0%; Score 13; DB 2; Length 1428;

Best Local Similarity 100.0%; Pred. No. 76;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 ggtgcagctcagc 25

Db 126 GGTGCAGCTCAGC 114

RESULT 39

; US-08-634-400-19/c

; Sequence 19, Application US/08634400

; Patent No. 5939068

; GENERAL INFORMATION:

; APPLICANT: BRAMS, Peter

; APPLICANT: CHAMAT, Soulaïma Salim

; APPLICANT: PAN, Li-Zhen

; APPLICANT: WALSH, Edward E.

; APPLICANT: HEARD, Cheryl Janne

; APPLICANT: NEWMAN, Roland Anthony

; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN

; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND

; NUMBER OF SEQUENCES: 19 METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Burns, Doane, Swecker & Mathis

; STREET: P.O. Box 1404

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 2213-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentln Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/634,400

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/488,376

; FILING DATE: 07-JUN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Teskin, Robin L.

; REGISTRATION NUMBER: 35,030

; REFERENCE/DOCKET NUMBER: 012712-150

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 836-6620

; TELEFAX: (703) 836-2021

; INFORMATION FOR SEQ ID NO: 19:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1428 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..1428

; US-08-634-400-19

Query Match 52.0%; Score 13; DB 2; Length 1428;

Best Local Similarity 100.0%; Pred. No. 76;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 agtgcagctcag 24

Db 127 AGTGCAGCTCAG 115

RESULT 40

; US-08-635-878-17/c

; Sequence 17, Application US/08635878

; Patent No. 5955364

; GENERAL INFORMATION:

; APPLICANT: BRAMS, Peter

; APPLICANT: CHAMAT, Soulaïma Salim

; APPLICANT: PAN, Li-Zhen

; APPLICANT: WALSH, Edward E.

; APPLICANT: HEARD, Cheryl Janne

; APPLICANT: NEWMAN, Roland Anthony

; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN

; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND

; NUMBER OF SEQUENCES: 19 METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Burns, Doane, Swecker & Mathis

STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/635,878
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-635-878-17

Query Match 52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 ggtgcaggtcagc 25
|||||
Db 126 ggtgcaggtcagc 114

RESULT 41
US-08-635-878-19/c
Sequence 19, Application US/08635878
Patent No. 5955364
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/635,878
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-635-878-19

Query Match 52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 agtgcaggtcag 24
|||||
Db 127 agtgcaggtcag 115

RESULT 42
US-08-770-057-17/c
Sequence 17, Application US/08770057
Patent No. 5958765
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,057
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030

REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-770-057-17

Query Match 52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 13 ggtgcaggtcagc 25
|||||
Db 126 ggtgcaggtcagc 114

RESULT 43
US-08-770-057-19/c
Sequence 19, Application US/08770057
Patent No. 5958765
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,057
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-770-057-19

Query Match 52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 12 aggtgcaggtcagc 24
|||||
Db 127 aggtgcaggtcagc 115

RESULT 44
US-09-335-697B-17/c
Sequence 17, Application US/09335697B
Patent No. 6200804
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/335,697B
FILING DATE: 06-Jul-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/770,057
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
SEQUENCE DESCRIPTION: SEQ ID NO: 17;
US-09-335-697B-17

Query Match 52.0%; Score 13; DB 4; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;

Matches 13: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Job time: 6973 sec

OY 13 ggtgcaggtcagc 25
|||||
DB 126 GGTGCAGGTACG 114

RESULT 45
US-09-335-697B-19/C

; Sequence 19, Application US/09335697B
; Patent No. 6200804

; GENERAL INFORMATION:

; APPLICANT: BRAMS, Peter

; CHAMAT, Soulaïma Salim

; PAN, Li-Zhen

; WALSH, Edward E.

; HEARD, Cheryl Janne

; NEWMAN, Roland Anthony

; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN

; MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND

; METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Burns, Doane, Swecker & Mathis

; STREET: P.O. Box 1404

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 2213-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/335,697B

; FILING DATE: 06-Jul-2000

; CLASSIFICATION: <unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/770,057

; FILING DATE: <unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Teskin, Robin L.

; REGISTRATION NUMBER: 35,030

; REFERENCE/DOCKET NUMBER: 012712-150

; TELEPHONE: (703) 836-6620

; TELEFAX: (703) 836-6620

; INFORMATION FOR SEQ ID NO: 19:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1428 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..1428

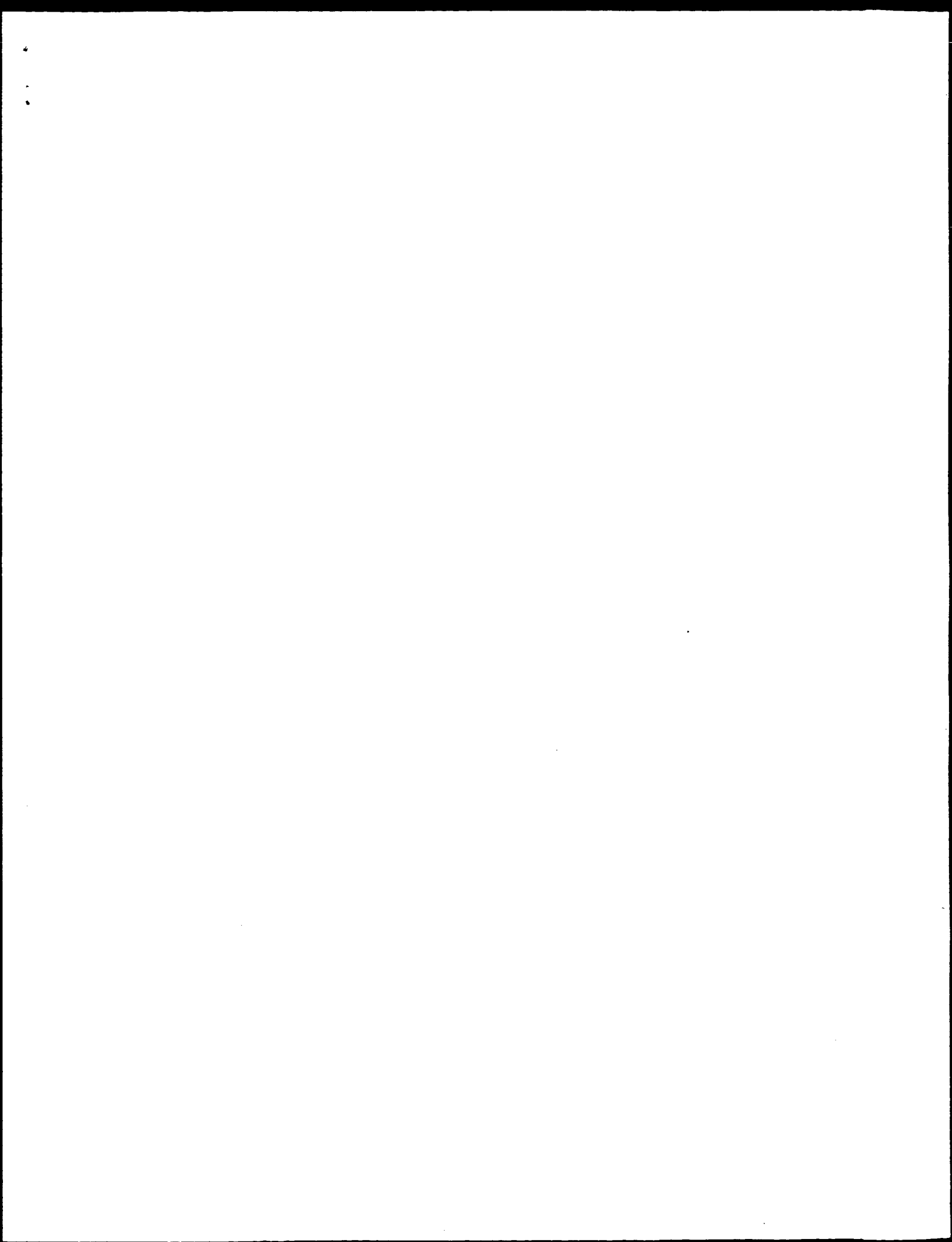
; SEQUENCE DESCRIPTION: SEQ ID NO: 19:

US-09-335-697B-19

Query Match 52.0%; Score 13; DB 4; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aggtgcaggtcag 24
|||||
DB 127 AGGTGCAGGTACG 115

Search completed: December 26, 2001, 12:55:41



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:10:35 ; Search time 4619.78 Seconds
(without alignments)
58.151 Million cell updates/sec

Title: US-09-396-196f-9

Perfect score: 25

Sequence: 1 cgatccctgcgcagtcgcagtcagc

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapect 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size: 9

Total number of hits satisfying chosen parameters: 544512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database:

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: qb_estl:*
11: qb_estl2:*
12: qb_hlc:*
13: qb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	68.0	324	11	BF464374
2	17	68.0	413	11	BE862745
3	17	68.0	495	10	A1789420
4	17	68.0	641	13	A2426129
5	17	68.0	767	11	BG703880
6	16	64.0	163	10	AA065941
7	16	64.0	288	10	BB516544
8	16	64.0	328	11	WS4113
9	16	64.0	338	11	W34110
10	16	64.0	368	10	AA110497
11	16	64.0	370	10	AA822106
12	16	64.0	430	10	A1467556

13	16	64.0	544	10	AW892602
14	16	64.0	553	13	A0530462
15	16	64.0	682	13	A2978539
16	16	64.0	775	10	BE535032
17	16	64.0	897	13	CNS021CM
18	16	64.0	976	11	BG172968
19	16	64.0	1011	12	AK010837
20	16	64.0	1298	11	BG328084
21	15	60.0	180	10	AA699872
22	15	60.0	264	11	H22139
23	15	60.0	279	11	R50061
24	15	60.0	293	13	AA877721
25	15	60.0	300	10	BB175346
26	15	60.0	306	11	BG100182
27	15	60.0	331	11	BT042497
28	15	60.0	345	10	AV633196
29	15	60.0	395	11	BF444068
30	15	60.0	396	10	AA486040
31	15	60.0	407	10	AA062258
32	15	60.0	407	11	BF599674
33	15	60.0	408	13	AA085001
34	15	60.0	409	11	BF653430
35	15	60.0	431	10	AA925187
36	15	60.0	454	11	BT042853
37	15	60.0	457	11	BF604067
38	15	60.0	466	11	BI398127
39	15	60.0	477	13	AO921326
40	15	60.0	492	11	BE808281
41	15	60.0	493	13	AE261270
42	15	60.0	496	11	BF652819
43	15	60.0	519	10	BE442332
44	15	60.0	543	13	A2168021
45	15	60.0	577	11	BG711675
46	15	60.0	607	13	A2873604
47	15	60.0	624	13	A2418216
48	15	60.0	639	10	AW342762
49	15	60.0	658	11	C98409
50	15	60.0	668	10	AU165518
51	15	60.0	678	13	AO361397
52	15	60.0	695	11	BG269235
53	15	60.0	710	13	A2988889
54	15	60.0	713	10	AU094890
55	15	60.0	743	10	BE309561
56	15	60.0	804	11	BF864835
57	15	60.0	841	10	BE376456
58	15	60.0	884	13	CNS05779
59	15	60.0	990	13	CNS024LL
60	15	60.0	1043	13	CNS05049
61	15	60.0	1064	11	BI078629
62	15	60.0	1069	13	CNS056GT
63	15	60.0	1126	11	BE795069
64	15	60.0	1143	11	BF124974
65	14	56.0	162	10	AV010258
66	14	56.0	196	10	AI210255
67	14	56.0	199	10	AI789574
68	14	56.0	204	11	BF772233
69	14	56.0	208	11	BF713183
70	14	56.0	218	11	BI016166
71	14	56.0	250	10	AV232136
72	14	56.0	257	11	BG408765
73	14	56.0	269	11	BF757622
74	14	56.0	274	13	AA065760
75	14	56.0	275	10	AV110186
76	14	56.0	279	10	AI097694
77	14	56.0	282	10	AI048306
78	14	56.0	283	10	AA322712
79	14	56.0	286	10	AA087394
80	14	56.0	288	11	BG382504
81	14	56.0	289	10	AI116568
82	14	56.0	291	11	BF994483
83	14	56.0	295	13	AO932051
84	14	56.0	300	10	AU099361
85	14	56.0	304	11	BF402359

AW892602	CN3-NN000
A0530462	RPCT-11-3
A2978539	2M0254K21
BE535032	601233227
BI175719	Tetraodon
BG172968	602336627
AK010837	Mus muscu
BG328084	602427134
AA699872	z181f04.s
H22139	y138a03.s1
R50061	y159c10.s1
AA877721	RPCT-23-1
BB175346	BI175346
BG100182	ux94g04.y
BT042497	RC5-OT009
AV633196	AV633196
BF444068	261798.MA
AA486040	69428.MAR
AA062258	m155908.r
BF599674	263546.MA
AA085001	HS-2269.A
BF653430	277257.MA
AA925187	WS1-76.E0
BT042853	RC5-OT009
BF604067	269755.MA
BI398127	UMN45505
AO921326	RPCT-23-2
BE808281	213447.MA
AE261270	RPCT-23-4
BF652819	276504.MA
BE442332	925017G02
A2168021	SP-0103.B
BG711675	Pg11n.-pk0
A2873604	2M0187J24
A2418216	1M0194P10
AW342762	f166h10.x
C98409	C98409
AU165518	AU165518
AO361397	mgXB0004A
BG269235	L0-3267T3
A2988889	2M0286J03
AU094890	AU094890
BE309561	601094957
BF864835	963055C06
BE376456	601262448
AL352782	Tetraodon
AL180930	Tetraodon
ALJ15090	Tetraodon
BI078629	602872730
AL336278	Tetraodon
BE795069	601592885
BF124974	601762117
AV010258	AV010258
AI210255	h4a02a1.r
AI789574	u126d08.r
BF772233	CM4-IT004
BF713183	MI-P-01-a
BI016166	PM1-ET025
AV232136	AV232136
BG408765	qB76c07.y
BF757622	CM4-CT057
AA065760	HS-2241.A
AV110186	AV110186
AI097694	ue35c01.y
AI048306	u68a002.y
AA322712	EST25645
AA087394	mol13b07.r
BG382504	298426.MA
AI116568	ud74h05.y
BF994483	CM3-GN010
AO932051	RPCT-23-2
AU099361	AU099361
BF402359	UI-R-CA0-

C 86	14	56.0	305	11	C72075	C72075	C72075	Nice
C 87	14	56.0	306	10	AA292999	z47f09.r	AA292999	
C 88	14	56.0	307	10	A1195641	A1195641	A1195641	
C 89	14	56.0	308	10	AM647049	ESR32551	AM647049	
C 90	14	56.0	311	10	AA135711	z02808.f	AA135711	
C 91	14	56.0	312	10	BR457497	BR457497	BR457497	
C 92	14	56.0	324	10	BC824822	602728925	BC824822	
C 93	14	56.0	327	10	BC824822	602728925	BC824822	
C 94	14	56.0	327	10	A1441334	s455d11.y	A1441334	
C 95	14	56.0	330	13	B52575	CIT-HSP-200	B52575	
C 96	14	56.0	335	13	A2762765	IM0557022	A2762765	
C 97	14	56.0	336	11	B1273545	F2_1a3	B1273545	
C 98	14	56.0	350	13	A2694363	AST-2HBC5	A2694363	
C 99	14	56.0	352	10	AM907954	ut87408.y	AM907954	
C 100	14	56.0	355	10	AM630381	hh82403.y	AM630381	
			361	10	AA672753	vs51d05.f	AA672753	

ALIGNMENTS

RESULT 1
LOCUS BE64374/c 324 bp mRNA EST 04-DEC-2000
DEFINITION UI-M-CG0P-bog-f-09-0-UI.1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone
VERSION BF64374
KEYWORDS BF64374.1 GI:11533557
SOURCE EST.
ORGANISM house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 324)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477

JOURNAL Contact: Chin, H
MEDLINE National Institute of Mental Health
COMMENT 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestr@mail.nih.gov

Oligo-dT track not found, Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares lab clone distribution: Researchers may obtain BMAP cDNA
clones from RESEARCH GENETICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements.
The following repetitive elements were found in this cDNA sequence:
102-169, >(GAA)n#Simple.Repeat
Seq primer: M13 Forward
POLYA=No.

FEATURES
source

Location/Qualifiers
1..324
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-CG0P-bog-f-09-0-UI"
/clone_lib="NIH_BMAP_Ret4_S2"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia)"
polylinker: Site_1: Not 1; Site_2: Eco RI; The
NIH_BMAP_Ret4_S2 library is a subtracted library,
ultimately derived from mouse retina tissue libraries at
various stages of development. For a detailed description
of the library from which this clone was derived, please
visit our web site at brainest.eng.uiowa.edu.
TAG_SFC=None found"
BASE COUNT 91 a 84 c 89 g 60 t

ORIGIN

Query Match 68.0%; Score 17; DB 11; Length 324;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 cgtcagtcgagtcag 24
|||||
Db 324 CGTCAGTGCAGTCAG 308

RESULT 2

LOCUS BE62745/c 413 bp mRNA EST 29-SEP-2000
DEFINITION UI-M-BH0-ajf-b-11-0-UI.f1 NIH_BMAP_M.S1 Mus musculus cDNA clone
VERSION BE62745
KEYWORDS BE62745.1 GI:10382015
SOURCE EST.
ORGANISM house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 413)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477

JOURNAL Contact: Chin, H
MEDLINE National Institute of Mental Health
COMMENT 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestr@mail.nih.gov

cDNA Library Preparation: M.B. Soares lab clone distribution:
Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
should be noted that Bento Soares is generating a small number of
additional specialized non-redundant arrays of BMAP cDNAs whose
availability will be considered under appropriate and limited
collaborative arrangements
Seq primer: M13 Reverse.
Location/Qualifiers
1..413
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH0-ajf-b-11-0-UI"
/clone_lib="NIH_BMAP_M.S1"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia)"
polylinker: Site_1: Not 1; Site_2: Eco RI; The
NIH_BMAP_M.S1 library is a subtracted library derived from
a mixture of normalized libraries from ten regions of the
mouse brain (cerebellum, brain stems, olfactory bulbs,
hypothalamus, cortex, amygdala, basal ganglia, pineal
gland, striatum, hippocampus). The driver used for
subtraction consisted of a pool of 20,000 cDNA clones
obtained from non-normalized and normalized libraries of
these ten regions of the mouse brain."

FEATURES
source

BASE COUNT 92 a 125 c 112 g 84 t
ORIGIN

Query Match 68.0%; Score 17; DB 11; Length 413;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 cgtcagtcgagtcag 24
|||||

RESULT	3
LOCUS	A1789420/c
DEFINITION	uk99a06.y1 Sugano mouse kidney mklia Mus musculus cDNA clone IMAGE:2064658 5', mRNA sequence.
ACCESSION	A1789420
VERSION	A1789420.1
KEYWORDS	GI:5337136
SOURCE	EST.
ORGANISM	house mouse. Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 495)
AUTHORS	Marras,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Stepien,M., Theising,B., Allen,M., Bowers,Y., Person ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Rittler ,E., Kohn,S., Shih,T., Jackson,Y., Cardenas,W., McCann,R., Waterston,R. and Wilson,R. The WashU-NCI Mouse EST Project 1999 Unpublished (1999) Other ESTs: uk99a06.x1
TITLE	Contact: Marras M/WashU-NCI Mouse EST Project 1999
JOURNAL	Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel.: 314 286 1800 Fax: 314 286 1810
COMMENT	Email: mouseest@wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:993846 Seq primer: custom primer used High quality sequence stop: 493.
FEATURES	Location/Qualifiers 1..495 /organism="Mus musculus" /strain="C57BL" /db_xref="taxon:10090" /clone="IMAGE:2064658" /clone_lib="Sugano mouse kidney mklia" /sex="female" /dev_stage="adult" /lab_host="DH10B" /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII (CACGCTCG); Site_2: DraIII (CACCAATGCG); 1st strand cDNA was primed with an oligo(dT) primer (ATGTGGCTTTTTTTTTTTTTTTT) double-stranded cDNA was ligated to a DraIII adaptor (TTGTGGCCCTACTGG), digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACGCTGCG, 3' site CACCAATGCG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCGTCTCTTAAGACCTGCG and 3' end primer CGACCTGCAGCTGCAGACA."
BASE COUNT	147 a 122 c 141 g 85 t
ORIGIN	
Query Match	68.0%; Score 17; DB 10; Length 495;
Best Local Similarity	100.0%; Pred. No. 44;
Matches	17; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
Oy	8 cgctcaggtcgacgctcag 24
Db	393 CGTCAGCTGCAGCTCAG 377
RESULT	4
A2426129/c	

LOCUS	AZ426129	641 bp	DNA	GSS	03-OCT-2000
DEFINITION	IM0206H03R Mouse 10kb plasmid UUGCM library Mus musculus genomic				
ACCESSION	clone UUGCM0206H03 R, DNA sequence.				
VERSION	AZ426129				
KEYWORDS	A2426129..1 GI:10550142				
SOURCE	GSS.				
ORGANISM	house mouse.				
	Mus musculus				
REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scurionath; Muridae; Murine; Mus.				
AUTHORS	1 (bases 1 to 641)				
	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,				
	Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Petersen,T., Rilly				
	,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.				
	and Wright,D., Weiss,R.				
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb				
	plasmid inserts				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Robert B. Weiss				
	University of Utah Genome Center				
	University of Utah				
	Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT				
	84112, USA				
	Tel.: 801 585 5606				
	Fax: 801 585 7177				
	Email: ddunn@genetics.utah.edu				
	Insert Length: 10000 Std Error: 0.00				
	Plate: 0206 row: H column: 03				
	Seq primer: CACACAGCAACACGCTATGACC				
	Class: plasmid ends				
	High quality sequence stop: 641.				
FEATURES	Location/Qualifiers				
Source	1..641				

```

/clone_lib="Mouse 10kb plasmid U00GCM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (91147321141gb)AT29072.11). A copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      156 a      184 c      183 g      118 t
ORIGIN

Query Match      68.0%; Score 17; DB 13; Length 641;
      Match Local Similarity 100.0%; Pred. No. 44;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY      8 cgtcaggtcaggtcag 24
      |||||||||||||||
Db      212 CGTCAGGTGCGAGTCAAG 196

```

BG703880/c 767 bp mRNA EST 07-MAY-2001
 LOCUS 602686932F1 NIH_MGC_95 Homo sapiens CDNA clone IMAGE:4819446 5',
 DEFINITION mRNA sequence.
 ACCESION BG703880
 VERSION BG703880.1 GI:13976654
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 767)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs@mail.nih.gov
 Tissue Procurement: Miklos Palcovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
 Toshiyuki and Piero Carninci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: L1AM10722 Row: p Column: 07
 High quality sequence stop: 536.
 Location/Qualifiers
 1..767
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:4819446"
 /clone_lib="NIH_MGC_95"
 /issue_type="hippocampus"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pBluescriptPL (modified
 pBluescript KS+); Site: 1: BamHI; Site 2: SalI; XhoI (gtcga
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTV-3',
 size-selected for average insert size 2.5 kb and
 normalized to 10^5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NHGRI), National
 Institutes of Health). Note: this is a NIH-MGC Library."
 BASE COUNT 139 a 215 c 202 g 211 t
 ORIGIN
 Query Match 68.0%; Score 17; DB 11; Length 767;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 8 gctcaggtgcaggtcag 24
 ||||||||||||||||
 Db 649 GCTCAGGTGCAGGTGAC 633
 RESULT 6
 AA065941/c 163 bp mRNA EST 03-FEB-1997
 LOCUS m51a01.r1 Stralagene mouse testis (#937308) Mus musculus CDNA
 DEFINITION clone IMAGE:515496 5', mRNA sequence.
 ACCESION AA065941
 VERSION AA065941.1 GI:1562644
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 163)
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Scheinberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

Theising, B., Wylie, T., Lemon, G., Soares, B., Wilson, R. and
 Maleson, R.,
 The WashU-HM Mouse EST Project
 Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 WashU-HM Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:309344
 Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 144.
 Location/Qualifiers
 1..163
 /organism="Mus musculus"
 /strain="Inbred CD-1"
 /db_xref="taxon:10090"
 /clone_image="IMAGE:515496"
 /clone_lib="Stralagene mouse testis (#937308)"
 /sex="males"
 /issue_type="testis"
 /dev_stage="10-12 week old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: testis; Vector: pBluescript SK-; Site: 1:
 PcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
 Oligo dT. Average insert size: 1.0 Kb; Uni-XAP XR Vector:
 -5' adaptor sequence: 5' GAATTCGGGACGAG 3' -3' adaptor
 sequence: 5' CTCGAGTGTGTTTGTGTTTGT 3'."
 BASE COUNT 23 a 56 c 49 g 35 t
 ORIGIN
 Query Match 64.0%; Score 16; DB 10; Length 163;
 Best Local Similarity 100.0%; Pred. No. 14e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 9 gtcaggtgcaggtcag 21
 ||||||||||||||||
 Db 16 GTCAGGTGCAGGTGAC 1
 RESULT 7
 BB516544/c 288 bp mRNA EST 28-JUL-2000
 LOCUS BB516544 RIKEN full-length enriched, 16 days neonate heart Mus
 DEFINITION musculus CDNA clone D830016J11 3' similar to M76601 Mouse alpha
 cardiac myosin heavy chain mRNA, mRNA sequence.
 ACCESION BB516544
 VERSION BB516544.1 GI:9568002
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 288)
 AUTHORS Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci,
 P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
 Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
 Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
 Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,
 Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
 Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata,
 Y., Shigemoto, Y., Shitagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,
 Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomianga, N., Toya,
 T., Tsunoda, Y., Watanabe, S., Watanabe, S., Yamamura, T., Yamazaki, I.,
 Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino,
 M., Muramatsu, M. and Hayashizaki, Y.
 RIKEN Mouse ESTs (Kono, H., et al.)
 Unpublished (2000)

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome-gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagao, S., Sasaki
, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Thermolabile and thermolabile enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki
, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

FEATURES

source

Location/Qualifiers
1..288
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_id="D830016J11"
/clone_lib="RIKEN full-length enriched, 16 days neonate
heart"
/tissue_type="heart"
/dev_stage="16 days neonate"
/lab_host="DH10B"
/note="Site_1: Salt; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGACAGCGCGCGACACGAGCTTTTCTTTTCTTTT 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGACAGCTTCTCGAGTTAATTAATTATATATCCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT

83 a 104 c 57 g 44 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 16; DB 10; Length 288;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 cgtcaggtcaggtcagca 23
|||||

Db 188 CGTCAGGTGACAGTCA 173

RESULT 8

LOCUS W54143 328 bp mRNA EST 03-JUN-1996
DEFINITION md13c03.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
clone IMAGE:368260 5', mRNA sequence.

ACCESSION W54143
VERSION W54143.1 GI:1355214

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 328)
Marta, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE

JOURNAL

The WashU-HMI Mouse EST Project
Unpublished (1996)

COMMENT

Contact: Marta M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:229692
Seq primer: mob.FEGA+ET.

FEATURES

source

Location/Qualifiers
1..328
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_id="IMAGE:368260"
/clone_lib="Soares mouse embryo NBME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTCCCAATCTGAAGTGGAGCGCGCGGGAATTTTCTTTTCTTTTCTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT7T3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Felima Bonaldo."

BASE COUNT

81 a 66 c 85 g 96 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 16; DB 11; Length 328;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 tcaagtcaggtcagc 25
|||||

Db 86 TCAGGTGACAGTCA 101

RESULT 9

LOCUS W34110 338 bp mRNA EST 13-MAY-1996
DEFINITION mc57c06.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
clone IMAGE:352618 5', mRNA sequence.

ACCESSION W34110
VERSION W34110.1 GI:1316081

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 338)
Marta, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

TITLE The Washu-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Maria M/Mouse EST Project
Washu-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:224418

Putative full length read
Seq primer: mob.RECA+ET
High quality sequence stop: 324.

FEATURES
source

1. .338
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="IMAGE:352618"
/clone_lib="Soares mouse embryo NDM13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACCAATCGAAGTGGAGCGCGCGCGAATATTTTCTTTTCTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne
State Univ., from 2 1; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Hento Soares and
M.Fatima Bonaldo. "

BASE COUNT 86 a 65 c 84 g 103 t
ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 338;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tcaggtgcaggtcagc 25
|||||
Db 82 TCAGGTGCAGGTGACG 97

RESULT 10
AA110497/c

LOCUS AA110497 368 bp mRNA EST 03-FEB-1997
DEFINITION m162e03.x1 Striatagene mouse testis (#937308) Mus musculus cDNA
clone IMAGE:516604 5', mRNA sequence.

ACCESSION AA110497
VERSION AA110497.1 GI:1662274

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 368)
Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lagy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE The Washu-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Maria M/Mouse EST Project

Washu-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:310452
Seq primer: -28m13 rev1 ET from Amersham.

FEATURES
source

1. .368
/organism="Mus musculus"
/strain="inbred CD-1"
/db_xref="taxon:10090"
/clone_image="IMAGE:516604"
/clone_lib="Striatagene mouse testis (#937308)"
/sex="males"
/tissue_type="testis"
/dev_stage="10-12 week old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: testis; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
-5' adaptor sequence: 5' GAATTCGGCGACGACG 3' -3' adaptor
sequence: 5' CTCGAGTCTTTTCTTTTCTTTTCTTTT 3' "

BASE COUNT 75 a 115 c 106 g 72 t
ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 368;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 gtcaagtgcaagtcag 24
|||||
Db 21 GTCAGTGCAGGTGACG 6

RESULT 11
AW822106

LOCUS AW822106 370 bp mRNA EST 17-MAY-2000
DEFINITION ug13c09.x1 Ren Stubs mouse thymus Mus musculus cDNA clone
IMAGE:2802352 3', mRNA sequence.

ACCESSION AW822106
VERSION AW822106.1 GI:7915183

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 370)
Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person
B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter
B., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.

TITLE The Washu-HHMI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT Other_ESTS: ug13c09.y1
Contact: Maria M/Washu-HHMI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1041876
Seq primer: Primer name ambiguous
High quality sequence stop: 303.

FEATURES
source Location/Qualifiers
1. .370


```

/organism="Mus musculus"
/strain="C3H"
/db_xref="taxon:10090"
/clone="IMAGE:2802352"
/clone_id="Ren Studbs mouse thymus"
/sex="mixed"
/dev_stage="3 weeks"
/lab_host="DH10B"
/note="Organ: thymus; Vector: p7T73D-Pac; Site_1: NotI;
Site_2: PacI; 1st strand cDNA was primed with an oligo(OT)
primer; double-stranded cDNA was ligated using 5' linker
ggcgcgcctat and 3' linker aactgaagaactat. Library is
size-selected >2.5 kb and average insert size is 3.5 kb.
Clones were arrayed from primary plating; non-amplified
library constructed by X. Ren and L. Studbs (Lawrence
Livermore National Laboratory and DOE Joint Genome
Institute, 7000 East Ave, L-453, Livermore, CA 94550)."
BASE COUNT      125 a      57 c      75 g      113 t
ORIGIN

Query Match      64.0%; Score 16; DB 10; Length 370;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      9 gtcagtcagtcagc 24
|||||
Db      53 CTCAGCTCAGCTCAG 68

RESULT 12
A1467556      430 bp      mRNA      EST      09-MAR-1999
LOCUS      A1467556
DEFINITION      ve36909.x1 Soares_mammary_gland_NbMNG Mus musculus cDNA clone
IMAGE:820288.3', mRNA sequence.
ACCESSION      A1467556
VERSION      A1467556.1 GI:4320893
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
1 (bases 1 to 430)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (infoimage.lnl.gov) for further information.
MG1:488568
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
High quality sequence stop: 406.
Location/Qualifiers
1..430
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:820288"
/clone_id="Soares_mammary_gland_NbMNG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: p7T73D-Pac (Pharmacia
) with a modified polylinker; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(OT)
primer 15'
TGTTCACATCTGAGTGGAGCGCCCGCAATGTTTTCATTTTTCATTTTTCATTTT
T 31; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into

```

```

the Not I and Eco RI sites of the modified p7T73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."
BASE COUNT      136 a      102 c      85 g      107 t
ORIGIN

Query Match      64.0%; Score 16; DB 10; Length 430;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      10 tcaagtcagtcagc 25
|||||
Db      259 TCAGCTCAGCTCAGC 244

RESULT 13
AM892602      544 bp      mRNA      EST      24-MAY-2000
LOCUS      CM3-NN0004-100300-111-f07 NN0004 Homo sapiens cDNA, mRNA sequence.
DEFINITION      AM892602
ACCESSION      AM892602
VERSION      AM892602.1 GI:8056807
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 544)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Brites,M.R.,
Nagal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
2020263
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-CM3-NN0004-100
300-111-f07&ts=2000-03-10&td=1)
Seq primer: puc 18 forward
High quality sequence stop: 543.
Location/Qualifiers
1..544
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="NN0004"
/dev_stage="Adult"
/note="Organ: nervous normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      127 a      152 c      160 g      105 t
ORIGIN

Query Match      64.0%; Score 16; DB 10; Length 544;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 9 gtcaggtgcagtcag 24
 DB 383 gtcaggtgcagtcag 398

RESULT 14
 A0530462/c
 LOCUS A0530462/c
 DEFINITION RPCI-11-369F11.TU RPCI-11 Homo sapiens genomic clone RPCI-11-369F11
 , DNA sequence.
 ACCESSION A0530462
 VERSION A0530462.1 GI:4842505
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 553)
 AUTHORS Zhao,S., Adams,M.D., Niernan,W., Malek,J., de Jong,P. and Venter,J.C.
 TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
 JOURNAL Unpublished (1997)
 COMMENT other GSSs: RPCI-11-369F11.TV
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbe@tigr.org
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@edj.org.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html.
 Seq primer: SP6
 Class: BAC ends

FEATURES
 Source Location/Qualifiers
 1..553
 /organism="Homo sapiens"
 /db_xref="GDB:7641442"
 /db_xref="taxon:9606"
 /clone="RPCI-11-369F11"
 /clone_1ib="RPCI-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"

BASE COUNT 108 a 165 c 157 g 123 t
 ORIGIN

Query Match 64.0%; Score 16; DB 13; Length 553;
 Best Local Similarity 100.0%; Pred.No. 1.4e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 gtcaggtgcagtcag 24
 DB 204 gtcaggtgcagtcag 189

RESULT 15
 A2978539/c
 LOCUS A2978539 682 bp DNA
 DEFINITION ZM0254K21R Mouse 10kb plasmid U0GC2M library Mus musculus genomic
 clone U0GC2M0254K21 R. DNA sequence.
 ACCESSION A2978539
 VERSION A2978539.1 GI:13849766
 KEYWORDS GSS.
 SOURCE house mouse.

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 682)
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausen,A. and Wright,D., Weiss,R.
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0254 row: K column: 21
 Seq primer: CACACAGCAACACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 682.

FEATURES
 Source Location/Qualifiers
 1..682
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U0GC2M0254K21"
 /clone_1ib="Mouse 10kb plasmid U0GC2M library"
 /sex="Female"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (914732114[gb]|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 173 a 158 c 120 g 231 t
 ORIGIN

Query Match 64.0%; Score 16; DB 13; Length 682;
 Best Local Similarity 100.0%; Pred.No. 1.4e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 gtcaggtgcagtcag 24
 DB 55 gtcaggtgcagtcag 40

RESULT 16
 BE535032/c
 LOCUS BE535032/c 775 bp mRNA
 DEFINITION 601233327FL NCI_GMP_Mam6 Mus musculus cDNA clone IMAGE:3556935 5', mRNA sequence.
 ACCESSION BE535032
 VERSION BE535032.1 GI:9763677
 KEYWORDS EST.

```

SOURCE      house mouse.
ORGANISM    Mus musculus
REFERENCE   1 (bases 1 to 897)
AUTHORS     NIH-MGC http://mgc.ncl.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgabs-remail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LHAM8775 row: f column: 08
            High quality sequence stop: 225.

FEATURES
  source
    1..775
    /organism="Mus musculus"
    /strain="FVB/N"
    /db_xref="taxon:10090"
    /clone="IMAGE:3596935"
    /clone_lib="NCI CGAP Mam6"
    /sex="female, virgin"
    /tissue_type="infiltrating ductal carcinoma"
    /dev_stage="5 months"
    /lab_host="DH10B"
    /note="Organ: mammary; Vector: pCMV-SPORT6; site_1: SalI;
    site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
    Library constructed by Life Technologies, Investigator
    providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT  170 a 186 c 221 g 198 t

ORIGIN
Query Match      64.0%; Score 16; DB 10; Length 775;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 cgtcagtcagtcag 23
    |||||||
Db 201 CGTCAGTGCAGTCA 186

RESULT 17
CNS021CM 897 bp DNA GSS 12-MAY-2000
LOCUS     Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 225c12 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL176719.1 GI:7814776
VERSION   GSS; genome survey sequence.
KEYWORDS  Tetraodon nigroviridis.
SOURCE    Tetraodon nigroviridis.
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 897)
AUTHORS   Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C.,
            Bouneau, L., Billault, A., Quetier, F., Sautin, W., Bernot, A. and
            Weissenbach, J.
TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 897)
AUTHORS   Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
            Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F.,
            Sautin, W. and Weissenbach, J.

SOURCE      Human gene number estimate provided by genome wide analysis using
ORGANISM    Tetraodon nigroviridis DNA sequence
REFERENCE   3 (bases 1 to 897)
AUTHORS     Direct Submission
TITLE       Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
JOURNAL     This sequence is a single read and was generated as part of a large
COMMENT     scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.

FEATURES
  source
    1..897
    /organism="Tetraodon nigroviridis"
    /db_xref="taxon:99883"
    /clone="225c12"
    /clone_lib="G"
    /note="Genoscope sequence ID : C0AG225BB06LP1-end : T7"

BASE COUNT  118 a 284 c 294 g 182 t 19 others

ORIGIN
Query Match      64.0%; Score 16; DB 13; Length 897;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 gtcagtcagtcag 24
    |||||||
Db 16 GTCAGTGCAGTCA 31

RESULT 18
BG172968 976 bp mRNA EST 06-FEB-2001
LOCUS     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus,
DEFINITION 602336627F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4459758 5',
mRNA sequence.
ACCESSION BG172968
VERSION   BG172968.1 GI:12679671
KEYWORDS  EST.
SOURCE    house mouse.
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 976)
AUTHORS   NIH-MGC http://mgc.ncl.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgabs-remail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LHAM10260 row: e column: 07
            High quality sequence stop: 514.

FEATURES
  source
    1..976
    /organism="Mus musculus"
    /strain="FVB/N"
    /db_xref="taxon:10090"
    /clone="IMAGE:4459758"
    /clone_lib="NCI CGAP Mam1"
    /tissue_type="tumor, biopsy sample"
    /dev_stage="3 months, virgin"
    /lab_host="DH10B"
    /note="Organ: mammary; Vector: pCMV-SPORT6; site_1: SalI;
    site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
    Library constructed by Life Technologies, Investigator
    providing samples: Gilbert Smith, NIH"

BASE COUNT  255 a 250 c 163 g 308 t

```


BASE COUNT 316 a 498 c 357 g 127 t
ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 1298;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 tcttcagtgacggtc 22
|||||
Db 984 TCAGTGCAGTGCAGTGC 969

RESULT 21
AA699872/c 180 bp mRNA EST 19-DEC-1997
LOCUS Z181f04.s1 Soares fetal_liver_spleen_INFLS_S1 Homo sapiens cDNA
DEFINITION clone IMAGE:461311 3', mRNA sequence.
AA699872
VERSION AA699872.1 GI:2702835
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 180)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S., Kitzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R., and Wilson, R.
JOURNAL Unpublished (1997)
COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -40m13 fwd. RT from Amerstham
High quality sequence stop: 144.
Location/Qualifiers

FEATURES

source

1. 180
/organism="Homo sapiens"
/db_xref="GDB:3752334"
/db_xref="taxon:9606"
/clone="IMAGE:461311"
/clone_lib="Soares fetal_liver_spleen_INFLS_S1"
/sex="male"
/dev_stage="20 week post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT773D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; This is a subcloned version of the original Soares fetal liver spleen INFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACTGCAAGATTATTAAGACTTTTCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT773 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 47 a 57 c 36 g 40 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 180;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 11 caggtgcaggtcagc 25
|||||

Db 138 CAGGTGCAGTGCAGC 124

RESULT 22
H22139/c 264 bp mRNA EST 06-JUL-1995
LOCUS Y138a03.s1 Soares breast 3NBHst Homo sapiens cDNA clone
DEFINITION IMAGE:160492 3', mRNA sequence.
H22139
VERSION H22139.1 GI:890834
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 264)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisakis, E., Waterston, R., Williamson, A., Wohldmann, P., and Wilson, R.
JOURNAL The WashU-Merck EST Project
Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Insert Size: 685
High quality sequence stops: 241
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert length: 685 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 241.
Location/Qualifiers

FEATURES

source

1. 264
/organism="Homo sapiens"
/db_xref="GDB:574535"
/db_xref="taxon:9606"
/clone="IMAGE:160492"
/clone_lib="Soares breast 3NBHst"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: breast; Vector: pT773D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTACCAATCTGAGGTGGAGCGGCCCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT773 vector (Pharmacia). Library went through one round of normalization to a Cot = 20. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 87 a 47 c 50 g 73 t 7 others
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 264;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tcaggtgcaggtcag 24
|||||

Db 215 TCAGGTGCAGGTGCAG 201

RESULT 23
R50061/c 279 bp mRNA EST 18-MAY-1995
LOCUS Y159c10.s1 Soares breast 2NBHst Homo sapiens cDNA clone
DEFINITION

T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamana, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.
RIKEN Mouse ESTs (Konno, H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
url:http://genome.res.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Thermolabile and thermostable activation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.res.riken.go.jp) for further details.

FEATURES

SOURCE

Location/Qualifiers

1..300

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="A230056E21"

/clone_lib="RIKEN full-length enriched, adult male hypothalamus"

/sex="male"

/tissue_type="hypothalamus"

/dev_stage="adult"

/lab_host="DH10B"

/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'

GAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCCAGATTAATTAATTAATCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda

FLC 1. "

BASE COUNT 66 a 81 c 58 g 95 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 300;

Best Local Similarity 100.0%; Pred. No. 4.4e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tcaggtcagcagtcag 24
|||||

DB 72 TCAGGTGTCAGTCAG 58

RESULT 26
BG100182/c 306 bp mRNA EST 29-JAN-2001
LOCUS

DEFINITION
Accession
Version
Keywords
Source
Organism
un94g04.y1 McCarrey Eddy type B spermatogonia Mus musculus cDNA
clone IMAGE:3656262 5', mRNA sequence.
BG100182
GI:12595499

REFERENCE

1 (bases 1 to 306)
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Rifter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/Mashu-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:1418566

Seq primer: Primer name ambiguous.

FEATURES

SOURCE

Location/Qualifiers

1..306

/organism="Mus musculus"

/strain="CD-1"

/db_xref="taxon:10090"

/clone="IMAGE:3656262"

/clone_lib="McCarrey Eddy type B spermatogonia"

/sex="male"

/tissue_type="type B spermatogonia, pooled from multiple mice"

/dev_stage="8 day"

/lab_host="DH10B (phage-resistant)"

/note="Organ: testis; Vector: pBluescript SK+ (stratagene); Site_1: XhoI; Site_2: EcoRI; cDNA oligo df-primed [5'-(GA)10-ACCTAGCTGCGATGTTTCTTTTCTT-3'] and directionally cloned using 5' linkers 5'-AATTCGACGACG-3' and 5'-CTCGTCCG-3'. Size selection of >400bp material gives average insert size ranging from 1-2 kb. Library was mass excised (from lambda-unizap-XR) and resulting single-stranded phagemids were prepped and transformed into DH10B. Library contains 968 recombinants. References: J. Androl. 20:635-639 and Gene 25:263-269. Library constructed and donated by J. McCarrey, Ph.D. (Southwest Foundation for Biomedical Research, Dept. of Genetics); excision done by E.M. Eddy, Ph.D. (National Institutes of Health, National Institute of Environmental Health Sciences). Original lambda-based library is available through ATCC, catalog #63417."

BASE COUNT 62 a 82 c 93 g
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 306;

Best Local Similarity 100.0%; Pred. No. 4.4e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 gtcaggtcagtcag 23
|||||

DB 95 CTCAGGTGTCAGTCA 81

RESULT 27
B1042497 331 bp mRNA EST 14-JUN-2001
LOCUS
DEFINITION RC5-OT0098-220101-011-F11 OT0098 Homo sapiens cDNA, mRNA sequence.
ACCESSION B1042497

VERSION B1042497.1 GI:14449123
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 331)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H., Brustein, A., deoliveira, P.S., Bucher, P., Joneneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC5&t2=RC5-070098-220101-011-f11&t3=2001-01-22&t4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 59
 High quality sequence stop: 331.
 Location/Qualifiers
 1..331
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="OT0098"
 /dev_stage="Adult"
 /note="Organ: ovary; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 54 a 100 c 111 g 66 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 11; Length 331;
 Best Local Similarity 100.0%; Pred. No. 4.4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 3 atccgcgcagctgc 17
 ||||||||||||||||
 Db 70 ATCTCGTCAGGTGC 84
 RESULT 28
 AV633196 345 bp mRNA EST 15-DEC-2000
 LOCUS AV633196 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
 DEFINITION cDNA clone HC017h07_r 5', mRNA sequence.
 ACCESSION AV633196
 VERSION AV633196.1 GI:10776516
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii.
 ORGANISM Chlamydomonas reinhardtii.
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.
 REFERENCE 1 (bases 1 to 345)
 Asamizu, E., Miura, K., Kuchio, K., Inoue, Y., Fukuzawa, H., Ohyama, K., Nakamura, Y. and Tabata, S.

TITLE Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii
 JOURNAL DNA Res. 7 (5), 305-307 (2000)
 MEDLINE 20539644
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizukazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
 FEATURES
 source
 1..345
 /organism="Chlamydomonas reinhardtii"
 /strain="C9"
 /db_xref="taxon:3055"
 /clone_lib="HC017h07_r"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI. The cDNA library was constructed from cells cultured in a medium with bubbling air containing 5% carbon dioxide"
 BASE COUNT 62 a 105 c 98 g 79 t 1 others
 ORIGIN
 Query Match 60.0%; Score 15; DB 10; Length 345;
 Best Local Similarity 100.0%; Pred. No. 4.4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 11 caggtgcagctcagc 25
 ||||||||||||||||
 Db 173 CAGTGCAGGTGACG 159
 RESULT 29
 BF444068 395 bp mRNA EST 01-DEC-2000
 LOCUS BF444068 261798 MARC 2Pig Sus scrofa cDNA 5', mRNA sequence.
 DEFINITION BF444068
 ACCESSION BF444068.1 GI:11504160
 VERSION EST.
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 REFERENCE 1 (bases 1 to 395)
 Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W. and Keele, J.W.
 Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
 Unpublished (2000)
 TITLE
 JOURNAL
 COMMENT Contact: Smith TP
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@mail.marc.usda.gov
 Single pass sequencing. Bases called and att_trimmed with phred v0.980904.e. Vector identified by cross_match with the -mnscore 18 and -minmatch 12 options.
 PCR Primers
 FORWARD: AGGAAACAGCTATGACCAT
 BACKWARD: GTTTCACGACGACGACG
 Plate: 96 row: D column: 1
 Seq primer: ATTTCGTCAGACATATAC.
 FEATURES
 source
 1..395
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /clone_lib="MARC 2Pig"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;

REFERENCE
AUTHORS
Bovidae: Bovinae: Bos.
1 (bases 1 to 407)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,
G.L., Heaton,M.P., Laegreid,W.M., Rohrer,G.A., Chitko-Mckown,C.G.,
Perte,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keeler,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 35 row: E column: 3
Seq primer: ATTAGTCACACTATAG.
Location/Qualifiers
1..407
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."
BASE COUNT 69 a 139 c 135 g 64 t
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 407;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
0Y 11 caggtgcaggtcagc 25
|||||
Db 93 CAGGTGCAGGTGACG 79

RESULT 33
AC085001/c
LOCUS
DEFINITION
Accession
VERSION
KEYWORDS
SOURCE
ORGANISM
Human sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 408)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA

FEATURES
source
1..408
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Plate=2269 Col=20 Row=C"
/sex="male"
/note="Organ: sperm; Vector: pHElobAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT 104 a 111 c 78 g 115 t
ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 408;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
0Y 11 caggtgcaggtcagc 25
|||||
Db 396 CAGGTGCAGGTGACG 382

RESULT 34
BF653430/c
LOCUS
DEFINITION
Accession
VERSION
KEYWORDS
SOURCE
ORGANISM
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 409)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,
G.L., Heaton,M.P., Laegreid,W.M., Rohrer,G.A., Chitko-Mckown,C.G.,
Perte,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keeler,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 69 row: P column: 10
Seq primer: ATTAGTCACACTATAG.
Location/Qualifiers
1..409
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; site_1: XbaI; site_2: XhoI;

FEATURES
source
1..409
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; site_1: XbaI; site_2: XhoI;

Library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle."

BASE COUNT 80 a 138 c 129 g 62 t

ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 409;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 cagctgcagctcagc 25
|||||
Db 255 CAGGTCAGTCAGC 241

RESULT 35
AM925187/c

LOCUS WS1_76_E02.B1.A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
DEFINITION mRNA sequence.

ACCESSION AM925187.1 GI:8091013

VERSION AM925187

KEYWORDS EST.

SOURCE Sorghum.

ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

REFERENCE 1 (bases 1 to 431)

AUTHORS Cordomier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
L.H.

TITLE An EST database from Sorghum: water-stressed plants

COMMENT Unpublished (2000)
Contact: Cordomier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 339
POLYA=No.

FEATURES
Location/Qualifiers
1..431

Source /organism="Sorghum bicolor"

/db_xref="taxon:4558"

/clone_lib="Water-stressed 1 (WS1)"

/note="Organ: Mix of 5-week old plants on days 7 & 8 after
water was withheld; Vector: Lambda Zap; Site_1: XhoI;
Site_2: EcoRI; The library was made from polyA RNA in the
cloning vector Lambda Zap II. Clones to be sequenced were
prepared by mass excision."

BASE COUNT 98 a 135 c 106 g 92 t

ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 431;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 gattcctgcagctcagtg 16
|||||
Db 419 GATCCTGCTCAGCTG 405

RESULT 36
BI042853/c

LOCUS BI042853 454 bp mRNA EST 14-JUN-2001

DEFINITION RC5-OT0098-020201-012-F07 OT0098 Homo sapiens cDNA, mRNA sequence.
ACCESSION BI042853
VERSION BI042853.1 GI:14449479

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 454)

REFERENCE Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL MEDLINE

COMMENT 20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?el=RC5&t2=RC5-OT0098-
020201-012-F07&t3=2001-02-02&t4=1)

Seq primer: puc 18 forward

High quality sequence stop: 128.

FEATURES
Location/Qualifiers
1..454

Source /organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="OT0098"

/dev_stage="Adult"

/note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORFESTS PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 97 a 140 c 132 g 81 t 4 others

ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 454;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 atctcgtcagctcagc 17
|||||
Db 427 ATCCTGCTCAGCTGC 413

RESULT 37
BF604067/c

LOCUS BF604067 457 bp mRNA EST 25-APR-2001

DEFINITION BF604067 MARC 380V Bos taurus cDNA 5', mRNA sequence.

ACCESSION BF604067

VERSION BF604067.1 GI:11702307

KEYWORDS EST.

SOURCE cow.

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 457)

Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,

TITLE
JOURNAL
MEDLINE
COMMENT

Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennet, G.L., Heaton, M.P., Laegreid, W.M., Rohrer, G.A., Chitko-McKown, C.G., Perte, G., Holt, L., Karameycheva, S., Liang, F., Quackenbush, J., and Keele, J.W.

Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

Genome Res. 11 (4), 626-630 (2001)

21180013

Contact: Smith 'PPL

USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smithemall.marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.

PCR Primers

FORWARD: AGGAACAGCATGACCAT
BACKWARD: GTTTCAGTCAGCAGC

Plate: 50 row: 1 column: 16

Seq primer: ATTTCAGTCAGCATATAG.

Location/Qualifiers

1..457

/organism="Bos taurus"
/db_xref="taxon:913"
/clone_lib="MARC 3BOY"
/tissue_type="pooled"
/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle."

BASE COUNT 82 a 152 c 142 g 80 t 1 others

ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 457;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caggtgcaggtcagc 25
|||||

DB 215 CAGGTGCAGTCAGC 201

RESULT 38
BI398127/c 466 bp mRNA EST 09-AUG-2001

LOCUS
DEFINITION UMR45605 Canine Brain cDNA Library Canis familiaris cDNA 5', mRNA sequence.

ACCESSION
BI398127
BI398127.1 GI:15145202

VERSION
EST.

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 466)
Roberts, M.C., Hendrickson, J.A., Hoffmann, D.E., Flickinger, G.H., Rutherford, M.S., and Mickelson, J.R.
University of Minnesota Canine Brain EST Project
Unpublished (2001)
Contact: Mickelson, J.R.
Veterinary Pathobiology
University of Minnesota
1988 Fitch Avenue, University of Minnesota, St. Paul, MN 55108, USA
Tel: 612 624 1246
Fax: 612 625 0204
Email: mickel001@umn.edu
Seq primer: M13 Reverse.
Location/Qualifiers
1..466

FEATURES
source

BASE COUNT 112 a 94 c 190 g 66 t 4 others

ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 466;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 gtcaggtcaggtcagc 23
|||||

DB 446 GTCAGTCAGTCAGTC 432

RESULT 39
AQ921326/c 477 bp DNA GSS 21-DEC-1999

LOCUS
DEFINITION RPCI-23-273P15.TV RPCI-23 Mus musculus genomic clone RPCI-23-273P15 DNA sequence.

ACCESSION
AQ921326
AQ921326.1 GI:6610329

VERSION
GSS.

KEYWORDS
SOURCE
ORGANISM

Mus musculus.
house mouse.
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 477)
Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akiret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P., and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@tigr.org, med.bufile@tigr.org). Clones may be purchased from BACPAC Resources (<http://bacpac.med.bufile.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
Plate: 273 row: P column: 15
Seq primer: T7
Class: BAC ends.

Location/Qualifiers

1..477

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="RPCI-23-273P15"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Site selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 101 a 107 c 125 g 144 t

ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 477;
 Best Local Similarity 100.0%; Pred. No. 4.4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 gtcaggtcaggtca 23
 |||||||
 Db 68 GTCAGGTGACGTCA 54

RESULT 40
 BE808281/c 492 bp mRNA EST 25-APR-2001
 LOCUS
 DEFINITION BE808281 213447 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION BE808281
 VERSION BE808281.1 GI:10239393
 KEYWORDS EST.
 SOURCE
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.

REFERENCE
 1 (bases 1 to 492)
 Smith,T.P., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,
 G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
 Perlea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
 Keefe,J.W.

TITLE
 Sequence evaluation of four pooled-tissue normalized bovine CDNA
 libraries and construction of a gene index for cattle
 Genome Res. 11 (4), 626-630 (2001)
 21180013

COMMENT

USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@maill.marc.usda.gov
 Single pass sequencing. Bases called and alt_trimmed with phred
 v0.980904.e. Vector identified by cross-match with the minscore 18
 and -minmatch 12 options.

PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCTCAGTACGACG
 Plate: 68 row: p column: 2
 Seq primer: ATTGAGTACACTATAG.

FEATURES

SOURCE

1. 492
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 2BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 library made from pooled tissue from testis, thymus,
 semitendinosus muscle, longissimus muscle, pancreas,
 adrenal, and endometrium."
 BASE COUNT 96 a 161 c 149 g 85 t 1 others
 ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 492;
 Best Local Similarity 100.0%; Pred. No. 4.4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caggtcaggtcaggtca 25
 |||||||
 Db 263 CAGGTGACGTCA 249

RESULT 41

A2261270/c

A2261270 495 bp DNA GSS 26-JUL-2000
 DEFINITION RPCI-23-44018.TV RPCI-23 Mus musculus genomic clone RPCI-23-44018.
 DNA sequence.
 ACCESSION A2261270
 VERSION A2261270.1 GI:9469499
 KEYWORDS GSS.

SOURCE

house mouse.
 Mus musculus

ORGANISM

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
 1 (bases 1 to 495)

AUTHORS

Zhao,S., Nierman,W., Feldlyum,T., Malek,J., Shatsman,S., Akiret,
 B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
 and Fraser,C.M.

TITLE

Mouse BAC End Sequences from Library RPCI-23
 unpublished (1999)
 Other GSSs: RPCI-23-44018.TJ

JOURNAL

Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

COMMENT

Email: szhae@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.html>)
 or from Resea ch Genetics (http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)
 Plate: 440 row: I column: 8
 Seq primer: T7
 Class: BAC ends.

FEATURES

SOURCE

1. 495
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-44018"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:
 EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methylase. Size
 selected DNA was cloned into the pBAC3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."
 BASE COUNT 106 a 110 c 132 g 147 t
 ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 495;
 Best Local Similarity 100.0%; Pred. No. 4.4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 gtcaggtcaggtca 23
 |||||||
 Db 66 GTCAGGTGACGTCA 52

RESULT 42

LOCUS

BF652819 496 bp mRNA EST 25-APR-2001

DEFINITION

276504 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.

ACCESSION

BF652819

VERSION

BF652819.1 GI:11917951

KEYWORDS

EST.

SOURCE

cow.

ORGANISM

Bos taurus

Eukaryota;

Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS

Mammalia; Euteria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 496)
Smith, T.P., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett,
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-Mokom, C.G.,
Perta, G., Holt, I., Karayancheva, S., Liang, F., Quackenbush, J. and
Keeler, J.W.

TITLE

Sequence evaluation of four pooled-tissue normalized bovine CDNA
libraries and construction of a gene index for cattle

JOURNAL
MEDLINE
COMMENT

21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smitht@mail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.

PCR Primers

FORWARD: AGGAACAGCATGACCAT
BACKWARD: GTTTCGCCAGTCACGACG
Plate: 67 row: C column: 19
Seq primer: ATTGAGTGACACTATAG.

FEATURES

source

Location/Qualifiers
1..496

BASE COUNT
ORIGIN

100 a 164 c 148 g 84 t

Query Match 60.0%; Score 15; DB 11; Length 496;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caggtgcaggtcagc 25
|||||

Db 290 CAGGTGCAGGTGACG 276

RESULT 43

BEA42332

LOCUS BEA42332 519 bp mRNA EST 25-JUL-2000
DEFINITION 925017G02.x1 C. reinhardtii CC-2290, normalized, lambda Zap II
Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION

BEA42332
BEA42332.1 GI:9441847

KEYWORDS

SOURCE

ORGANISM

Chlamydomonas reinhardtii.
Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.

1 (bases 1 to 519)
Grossman, A., Davies, J., Pederspiel, N., Harris, E., Iellevre, P.,
McDermott, J.P., Silflow, C., Stern, D. and Surzycki, R.

Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 2

Unpublished (2000)
Contact: Elizabeth H. Harris
DCMB Box 91000

JOURNAL
COMMENT

Duke University
Durham, NC 27708-1000, USA

FEATURES

source

Tel: 919 613 8164
Fax: 919 613 8177
Email: chlamy@duke.edu.

Location/Qualifiers

1..519

/organism="Chlamydomonas reinhardtii"

/strain="CC-2290 wild type mt- S1 D2"

/db_xref="taxon:3055"

/clone_lib="C. reinhardtii CC-2290, normalized, lambda Zap

II"

/note="Vector: Bluescript II SK-. Site_1: EcoRI; Site_2:

XhoI; This library was constructed by John Davies and

Jeffrey McDermott. RNA was isolated from strain CC-2290

(Minnesota isolate of C. reinhardtii) grown to mid-log

phase in TAP (acetate containing) medium in the light.

PolyA mRNA was purified, and cDNA was synthesized and

directionally cloned into lambda Zap II (Stratagene) in

the EcoRI (5') and XhoI (3') sites. Bluescript II SK-

plasmids were excised from the lambda Zap clones by

superinfection with ExAssist (Stratagene) phage. The

library was normalized using method 4 described in Bonaldo

et al (1996) Genome Research 6: 791-806."

BASE COUNT
ORIGIN

137 a 137 c 121 g 124 t

Query Match 60.0%; Score 15; DB 10; Length 519;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctcgtcaggtcagc 18
|||||

Db 72 TCCTGTCAGGTGCA 86

RESULT 44

A2168021/C

LOCUS A2168021 543 bp DNA GSS 29-AUG-2000
DEFINITION SP_0103.B1.P05.T7A Strongylocentrotus purpuratus, purple sea urchin
clone Plate=103 Col=9 Row=L, DNA sequence.

ACCESSION

A2168021
A2168021.1 GI:8338389

VERSION

KEYWORDS

SOURCE

GSS.
Strongylocentrotus purpuratus.

Strongylocentrotus purpuratus.
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Echinoidea; Echinacea; Echinoida;

Strongylocentrotidae; Strongylocentrotus.

1 (bases 1 to 543)
Cameron, R.A., Mahaffas, G., Rast, J.P., Martinez, P., Biondi, T.R.,
Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray,
G.A., Eftensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and
Hood, L.

A sea urchin genome project: Sequence scan, virtual map, and
additional resources

Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)

20402566

Contact: Cameron, RA, Davidson, EH, Hood, L

Division of Biology 156-29

California Institute of Technology

Pasadena California 91125, USA

Tel: (626) 395-8421

Fax: (626) 793-3047

Email: acameron@caltech.edu

Plate: 103 row: L column: 9

Seq primer: T7

Class: BAC ends

High quality sequence stop: 543.

FEATURES

source

Location/Qualifiers
1..543
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7568"

```

/clone="plate=103 Col=9 Row=L"
/clone_lib="Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli
DH10B"
BASE COUNT      149 a      116 c      116 g      157 t      5 others
ORIGIN

```

```

Query Match          60.0%; Score 15; DB 13; Length 543;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3 atccctcaggtgc 17
        |||
        115 ATCTCTGTCAGCTGC 101

```

```

RESULT 45
BG711675      577 bp      mRNA      EST      08-MAY-2001
LOCUS
DEFINITION
    pg11n.pk008.m3 Normalized Liver Library Gallus gallus cDNA clone
    pg11n.pk008.m3 5' similar to sp|Q91766|HNF4_XENLA HEPATOCYTE
    NUCLEAR FACTOR 4-ALPHA (HNF-4-ALPHA) emb|CA85763.1| (237526)
    hepatocyte nuclear factor 4 (HNF4) [Xenopus laevis]G, mRNA
    sequence.

```

```

ACCESSION      BG711675
VERSION        BG711675.1 GI:14005625
KEYWORDS
SOURCE
    chicken.
    Gallus gallus

```

```

ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
    Phasianinae; Gallus.
    1 (bases 1 to 577)
    Burnside,J., Morgan,R.W. and Cogburn,L.A.
    Chicken ESTs from a normalized liver library
    Unpublished (2001)
    Contact: Joan Burnside
    Molecular Endocrinology
    University of Delaware
    40 Townsend Hall, Newark, DE 19717, USA
    Tel: 302 831-1345
    Fax: 302-831-3411
    Email: joan@udel.edu, www.chickest.udel.edu.

```

```

FEATURES
    source
        1..577
            Location/Qualifiers
                /organism="Gallus gallus"
                /db_xref="taxon:9031"
                /clone="pg11n.pk008.m3"
                /clone_lib="Normalized Liver Library"
                /sex="Male and Female"
                /tissue_type="liver"
                /lab_host="E.coli EMDH10B"
                /note="Vector: PCMVSPORT 6"

```

```

BASE COUNT      125 a      165 c      164 g      121 t      2 others
ORIGIN

```

```

Query Match          60.0%; Score 15; DB 11; Length 577;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      11 caggtgcaggtgc 25
        |||
        335 CAGGTGCAGCTCAGC 349

```

```

Search completed: December 26, 2001, 14:10:41
Job time: 7088 sec

```

